Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2024

Supporting Information

Polymer Design via SHAP and Bayesian Machine Learning Optimizes pDNA and CRISPR Ribonucleoprotein Delivery

Rishad J. Dalal[‡], Felipe Oviedo[§], Michael C. Leyden[†], Theresa M. Reineke[‡*]

‡Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, United States
§Nanite Inc., Boston, Massachusetts 02109, United States
†Department of Chemical Engineering and Material Science, University of Minnesota, Minneapolis, Minnesota 55455, United States

Table of Contents

MATERIALS 4

CHEMICAL REAGENTS 4
POLYPLEX/TRANSFECTION REAGENTS 4
CELL CULTURE REAGENTS 4
SIZE EXCLUSION CHROMATOGRAPHY: 4
NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY 4

SYNTHESIS 5

PENTAFLUORO PHENYLMETHACRYLATE (PFPMA) MONOMER SYNTHESIS 5
BENZIMIDAZOLE ETHANETHIOL (BET) SYNTHESIS 7
SYNTHESIS OF POLYPENTAFLUOROPHENYL METHACRYLATE (PPFPMA) AND POST POLYMERIZATION MODIFICATION TO POLYALLYL METHACRYLAMIDE (PAMAM) 8
POST POLYMERIZATION FUNCTIONALIZATION THROUGH THIOL-ENE CLICK CHEMISTRY 9
HOMOPOLYMER FUNCTIONIZATION 9
TIME KINETICS OF THIOL-ENE CLICK 9
THIOL CONCENTRATION KINETICS OF THIOL-ENE CLICK 10
CO-POLYMER FUNCTIONALIZATION 11
OFFLINE BATCH MODE MEASUREMENT OF ∂N/∂C OF MACROMONOMER AND BOTTLEBRUSH 14
PK_A TITRATIONS OF THE FOUR THIOL SMALL MOLECULES 14
CLOGP VALUES 15

BIOLOGICAL AND PHYSICAL CHARACTERIZATION OF POLYPLEXES 17

CELL CULTURE PROCEDURES 17
GENERAL PDNA AND RNP POLYMER TRANSFECTION PROTOCOL 17
GENERAL FLOW WORKUP FOR QUANTIFICATION OF EXPRESSION 17
GENERAL CCK-8 VIABILITY ASSAY 17
DYNAMIC LIGHT SCATTERING (DLS) WITH PDNA 18
DYNAMIC LIGHT SCATTERING (DLS) WITH RNP 20
DYE EXCLUSION WITH PDNA 22
DYE EXCLUSION WITH RNP 24
ROUND 1 PDNA AND MCHERRY TRANSFECTION 25
27
ROUND 2 PDNA AND MCHERRY TRANSFECTION 31
ROUND 3 PDNA AND MCHERRY TRANSFECTION 34

ANALYSIS BY MACHINE LEARNING 35

POLYMER FEATURE ATTRIBUTION USING SHAP 35
BAYESIAN OPTIMIZATION (BO) 36

IN VIVO HYDRODYNAMIC TAIL VEIN 37

ANIMALS 37
REAGENTS 37
HYDRODYNAMIC INJECTIONS. 37
IN VIVO BIOLUMINESCENT IMAGING 38

REFERENCES 39

Materials

Chemical reagents

All solvents purchased were ACS grade. Dialysis tubing (Mw cut-off = 1 kDa, 40 kDa) was purchased from Spectra/Por, treated with 0.1 wt % ethylenediaminetetraacetic acid (EDTA) solution, and stored in a ~0.05 wt % sodium azide solution. The tubing was soaked and rinsed with Milli-Q water prior to use. The following were purchased from Millipore Sigma (Massachusetts, USA): cysteamine hydrochloride (>98%, titration), 2-(dimethylamino) ethanethiol hydrochloride (95%), 2-(diethylamino) ethanethiol hydrochloride (95%), allylamine (98%), triethylamine (99.5%), 4,4′-azobis(4-cyanovaleric acid) (V-501, >98%), 4-cyano-4-[(dodecylsulfanylthiocarbonyl) sulfanyl] pentanoic acid (97%, HPLC), methacryloyl chloride (97%), 2,2-dimethoxy-2-phenylacetophenone (DMPA, 99%), 3-mercaptopropionic acid (>99%), *o*-phenylenediamine (99.5%), 2,6-lutidine (98%). Pentafluorophenol (>99%) was purchased from Oakwood Chemical (South Carolina, USA). Hydrochloric acid (36.5-38%), ACS grade, was purchased from VWR (Pennsylvania, USA).

Polyplex/transfection reagents

pZsGreen (4.7 kb) was originally purchased from Clontech (Mountain View, CA) and amplified by Aldevron (Fargo, ND). spCas9 protein (162 kDa), containing 2 nuclear localization signals, was purchased from Aldevron. sgRNA was purchased from Synthego (CA, USA). A cell counting kit-8 (CCK-8) was purchased from Dojindo Molecular Technologies (Rockville, MD). Lipofectamine 2000, calcein violet stain, PicoGreen (Quant-iT PicoGreen, dsDNA reagent), and ultrapure ethidium bromide (10 mg/mL) were purchased from ThermoFisher Scientific (Waltham, MA).

Cell culture reagents

Dulbecco's Modified Eagle Medium (DMEM; high glucose, pyruvate, and glutamax supplemented), Fluorobrite DMEM (phenol red-free media), reduced serum medium (Opti-MEM), trypsin-EDTA (0.05%) with phenol red, phosphate buffered saline (PBS) pH = 7.4, ultrapure DNAse/RNAse-free distilled water (DI H_2O) antibiotic-antimycotic (100×), and heat inactivated fetal bovine serum (HI FBS) were purchased from Life Technologies ThermoFisher Scientific (Carlsbad, CA). The Mark Osborne laboratory at the University of Minnesota supplied the HEK293T cell line.

Size Exclusion Chromatography:

All pAMAm scaffold polymers were analyzed on a SEC-MALS instrument in DMF containing 0.05 M LiBr. All 36 polymers in the library were analyzed on a SEC-MALS instrument in H_2O containing 0.1 M Na_2SO_4, 1% acetic acid. Approximately 2-4 mg of sample was dissolved in 1 mL of the mobile phase and filtered through a 0.2 μm PTFE (for DMF) or GDHP (for H_2O) membrane filter before introduction into an Agilent Infinity 1200 HPLC system operating at 50 °C and 1.0 mL/min. The instrument was equipped with a Wyatt Optilab T-rEX differential refractive index detector and a Wyatt HELEOS-II MALS detector. Data analysis was performed using ASTRA® software.

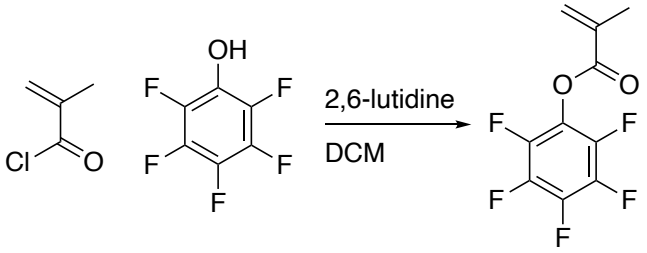
Nuclear Magnetic Resonance Spectroscopy

^{1}H NMR and ^{19}F NMR spectroscopy experiments were performed using a Bruker Advance III HD 500 spectrometer equipped with a 5 mm Prodigy TCI cryoprobe with z-axis gradients at

22°C using a 10 second relaxation delay and at least 32 transients without spinning to reduce signal-to-noise ratio in D_2O. ^1H NMR data was processed with Bruker TopSpin 3.5 pl 7 and MestReNova software.

Synthesis

Pentafluoro phenylmethacrylate (PFPMA) monomer synthesis



Scheme S1. Esterification of methacryloyl chloride with pentafluoro phenol to make the PFPMA monomer.

Pentafluorophenol (33 g, 0.179 mol) was dissolved in 100 mL anhydrous DCM and cooled in an ice bath for 30 min. After cooled to 0 °C, 2,6-lutidine (20.8 mL, 0.179 mol) was added and stirred over ice, followed by an additional 50 mL of anhydrous DCM. Methacryloyl chloride (21 mL, 0.215 mol) was mixed with 50 mL anhydrous DCM and slowly added to a round-bottom flask via an addition funnel over a span of 1 h. When the addition was complete, 2,6-lutidine-HCl, a white solid, crashed out of the solution. The reaction was stirred overnight at room temperature. The next day, the reaction mixture was filtered to remove the white salt, and then an aqueous work up was completed with DI water (3x200 mL) and brine (1x150 mL). The aqueous later was further extracted with DCM (1x200 mL). The reaction mixture was dried over magnesium sulfate, filtered, and then concentrated via *vacuo*. The slight yellow liquid product was further purified via vacuum distillation at 40 °C and 100 mTorr to yield a colorless non-viscous liquid of PFPMA (36 g, 80 %). ^1H NMR (400 MHz, Chloroform-*d*) δ 6.45 (s, 1H), δ 5.91 (s, 1H), δ 2.09 (s, 3H). ^{19}F NMR (400 MHz, Chloroform-*d*) δ -152.8 (d, 2 F), -158.2 (t, 1 F), -162.5 (d, 2 F).

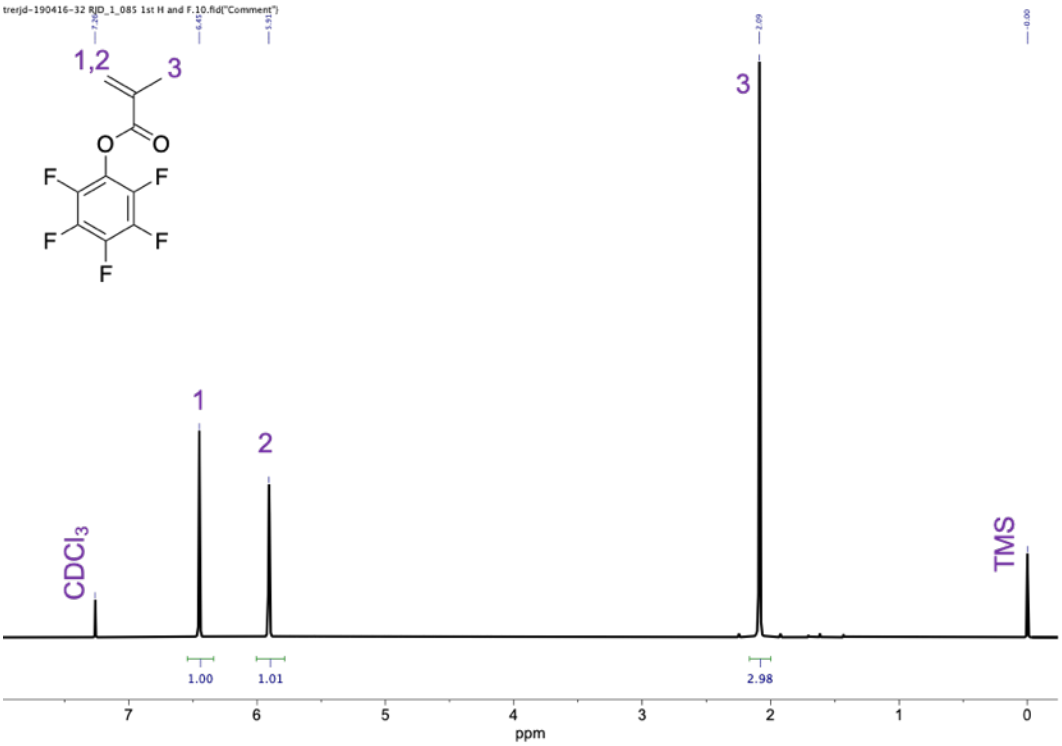


Figure S1: ^{1}H NMR spectrum of PFPMA monomer.

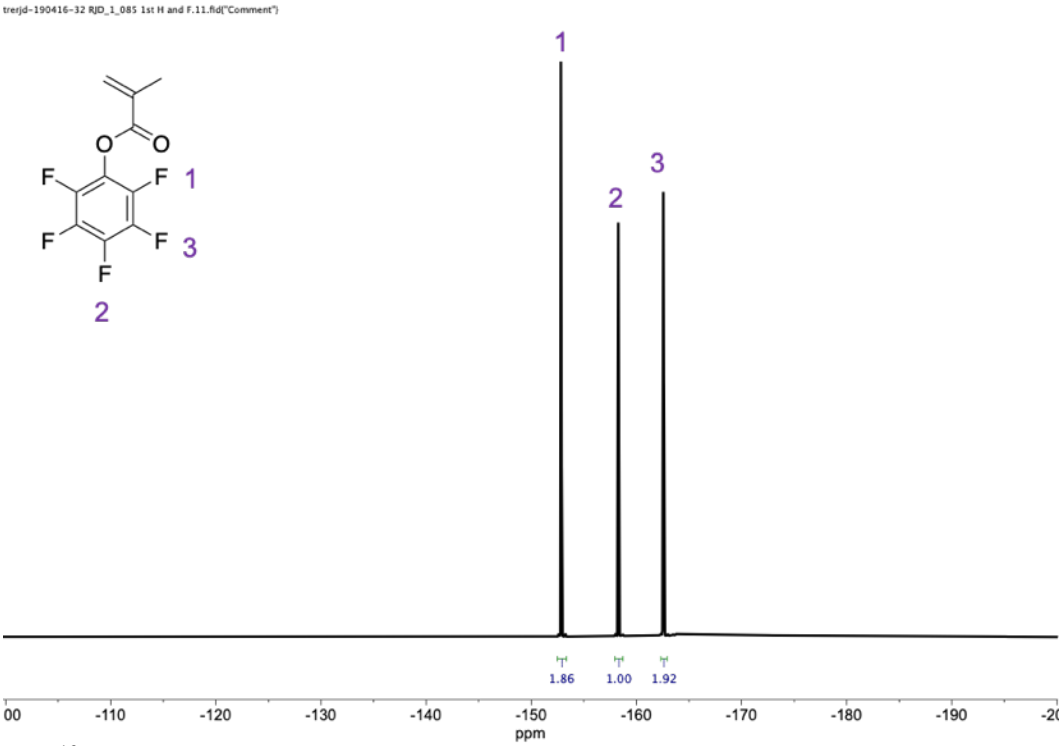
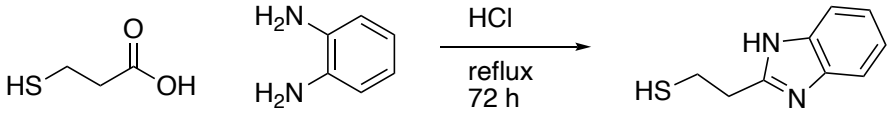


Figure S2: ^{19}F NMR spectrum of PFPMA monomer.

Benzimidazole ethanethiol (BET) synthesis



Scheme S2. Double condensation reaction of *o*-phenylenediamine with 3-mercaptopropionic acid to result in benzimidazole ethanethiol (BET).

3-mercaptopropionic acid (3.214 mL, 36.9 mmol) was added to *o*-phenylenediamine (3.285g, 30.4 mmol) and refluxed in HCl (4 M, 15 mL) for 72 h. The mixture started out a deep orange/brown color and transitioned to a dark green. The solubility of the mixture also increased as the reaction continued. The crude reaction was removed from the heat source and diluted in a beaker with 200 mL nanopore H_2O. A 50/50 wt% NaOH solution was added dropwise to the beaker and white precipitates formed. Precipitates were filtered and freeze dried. The result was an off-white crystalline solid of benzimidazole ethanethiol (BET) (2.3 g, 42%). ^{1}H NMR (400 MHz, MeOD) δ 7.48 (nfom, 2 H), δ 7.19 (nfom, 2H), δ 3.16 (t, 2 H), δ 2.94 (t, 2 H). Followed literature precedence.[1]

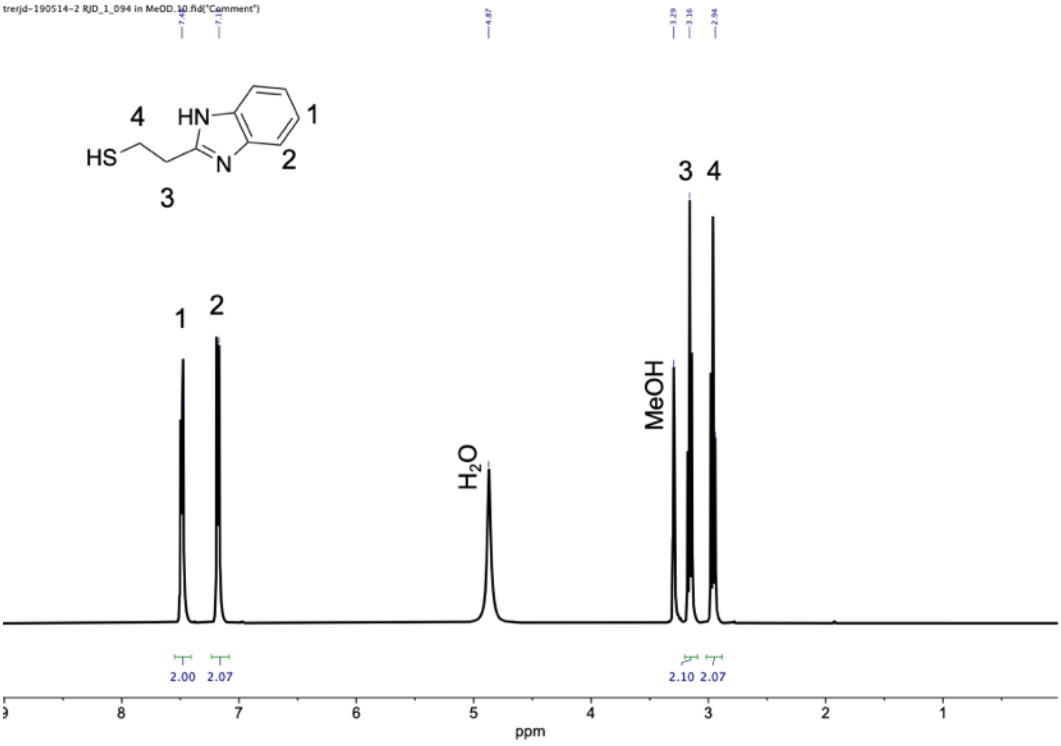
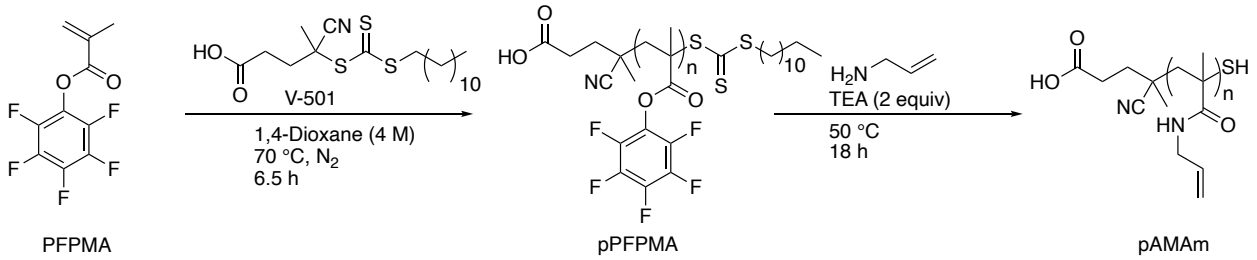


Figure S3: ^{1}H NMR spectrum of BET product.

Synthesis of polypentafluorophenyl methacrylate (pPFPMA) and post polymerization modification to polyallyl methacrylamide (pAMAm)



Scheme S3. RAFT polymerization of PFPMA and post polymerization modification to produce pAMAm.

Three polymerizations were done with varying equivalence on monomer (20, 50, and 100 equiv.) while keeping the CTA: initiator ratios used set (1:0.05) to produce three polymer lengths. The monomer of PFPMA was mixed with 4-cyano-4-[(dodecylsulfanylthiocarbonyl) sulfanyl] pentanoic acid, and V-501 in 1,4-dioxane (4 M). The solution was degassed for 40 min via nitrogen purging before heating to 70 °C for 6.5 h under nitrogen positive pressure. Polymerization was quenched by cooling the reaction in liquid nitrogen and exposure to atmosphere. An aliquot was precipitated into pentanes and then filtered for pPFPMA NMR (^{1}H and ^{19}F) and ATR-FTIR analysis. ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.25 (d, 2H), 1.52 – 0.99 (m, 3H). ^{19}F NMR (400 MHz, Chloroform-*d*) δ -150.80 (2 F), -156.99 (1 F), -161.86 (2 F). ATR-FTIR: 1778 cm^{-1} (C=O, ester), 1518 cm^{-1} (C=C, aromatic), 998 cm^{-1} (C-O), 989 cm^{-1} (C-O). The remaining pPFPMA was immediately reacted with a solution of allylamine and triethylamine mixture (2 equiv. to PFPMA monomer) through dropwise addition over 10 min. Solution was stirred at 50 °C under inert atmosphere for 16 hours. Reaction was quenched by cooling the reaction in ice. The obtained polymer was diluted and purified in a 1 kDa dialysis bag in MeOH and further concentrated via *vacuo* to obtain a clear flaky yellowish-brown solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.69 (s, 1H), 5.76 (s, 1H), 5.04 (s, 2H), 3.88 (d, 2H), 1.87 (s, 1H), 1.26 (s, 5H). ATR-FTIR: 1662 cm^{-1} (C=C, allyl), 1632 cm^{-1} (C=O, amide). Polymers were analyzed using DMF SEC-MALLS (0.05 M LiBr). The ∂n/∂c values were calculated by 100% mass recovery and ranged between 0.0969 and 0.1030 depending on polymer size.

pAMAm characterization for statistical polymer scaffolds:

Short (Sh): M_n=11 kDa, M_w=14 kDa, N=90, $Đ$= 1.25
Medium (Md): M_n=24 kDa, M_w=30 kDa, N=190, $Đ$= 1.28
Long (Lg): M_n=31 kDa, M_w=43 kDa, N=250, $Đ$= 1.37

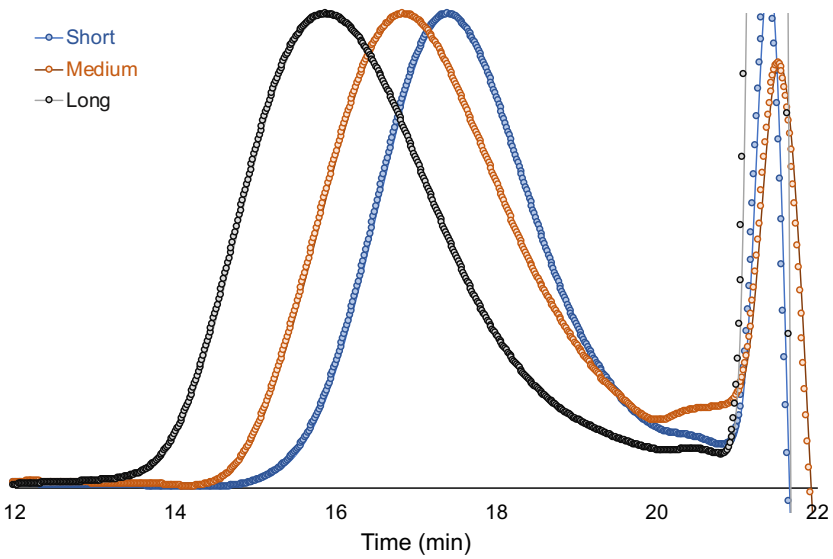
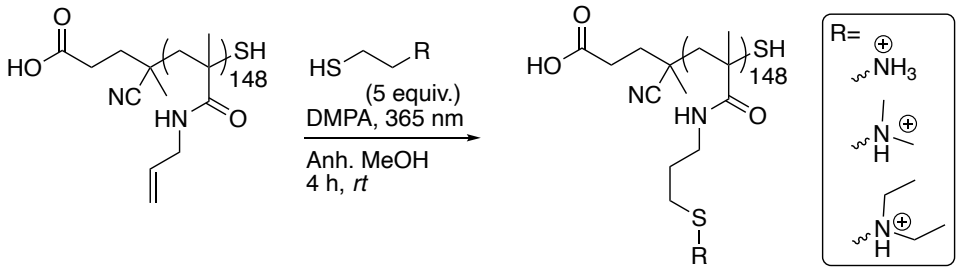


Figure S4: DMF SEC-dRI trace of the three pAMAm scaffolds.

Post polymerization functionalization through thiol-ene click chemistry

Homopolymer functionization



Scheme S4. Homo functionalization of pAMAm polymers

The three different lengths of pAMAm (ru: 90, 190, 250), 1 equiv., 4 mmol relative to monomeric form, 0.5 g) were dissolved in MeOH (1 M, 4 mL) and mixed with one of three thiol-amine salts, cysteamine hydrochloride, captamine hydrochloride, diethylamino ethanethiol hydrochloride (5 equiv., 2.0 mmol) and 2,2-dimethoxy-2-phenylacetophenone (DMPA) photoinitiator (0.05 equiv). The solutions ran in an open capped 20 mL scintillation vial with a stir bar at room temperature in a UV (~365 nm) gel nail curing box for 4 h. The resulting polymer solution was acidified with 0.5 mL HCl (1M) and then purified in a 1 kDa dialysis bag in millipore water and further concentrated via lyophilization to obtain a fluffy off-white powder. The cysteamine HCl functionalized polymer collapsed overtime to a glassy looking tacky solid. The mass of each polymer obtained was about 100-150 mg.

Time kinetics of thiol-ene click

The scaffold pAMAm polymer (1 equiv., 0.8 mmol, 0.1 g) was mixed with cysteamine HCl (1 equiv., 0.8 mmol, 0.091 g) and DMPA photoinitiator (5 mol% relative to thiol) in d-DMSO (0.1 M, 8 mL). The resulting stock solution was split into 10 samples and reacted under UV light. Samples were taken out at varying timepoints (time = 0, 0.25, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 4.0

h). To establish an internal standard, samples were spiked with a stock solution of phenol in d-DMSO. Timepoints were measured by ^{1}H NMR to monitor alkene depletion over the 4 h window and it was found that the reaction kinetics were complete after 0.5 h.

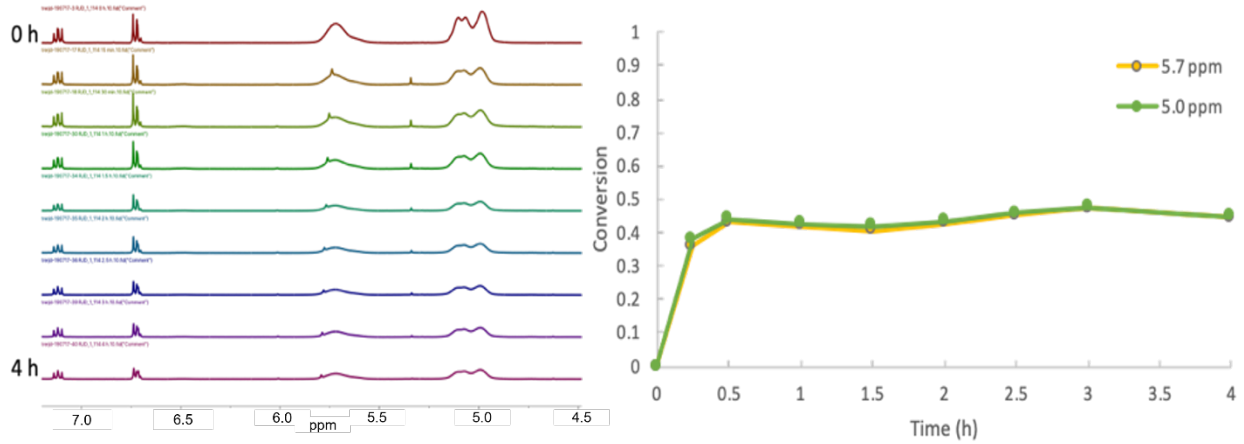


Figure S5. ^{1}H NMR kinetics showing conversion over time comparing the allyl alkene peaks at 5.7 and 5.0 ppm to an internal standard of phenol (7.1 and 6.7 ppm).

Thiol concentration kinetics of thiol-ene click

Due to incomplete functionalization at 1 equivalent, a series of reactions were conducted increasing the ratio of thiol to alkene. The scaffold pAMAm polymer (1 equiv., 0.08 mmol, 0.01 g) was mixed with cysteamine HCl at varying equivalence (0.5, 1.0, 1.5, 2.0, and 3.0 equiv.) and DMPA photoinitiator (5 mol% relative to thiol) in d-DMSO (0.1 M, 0.8 mL) and then reacted under UV light for 1 h. To establish an internal standard, samples were spiked with a stock solution of phenol in d-DMSO. By comparing to the phenol peak as well as the incorporation of the methylene peaks, ^{1}H NMR was used to measure timepoints and monitor both alkene depletion and the increase in thiol concentration. To achieve full conversion, 5 equiv. thiol was used in subsequent experiments.

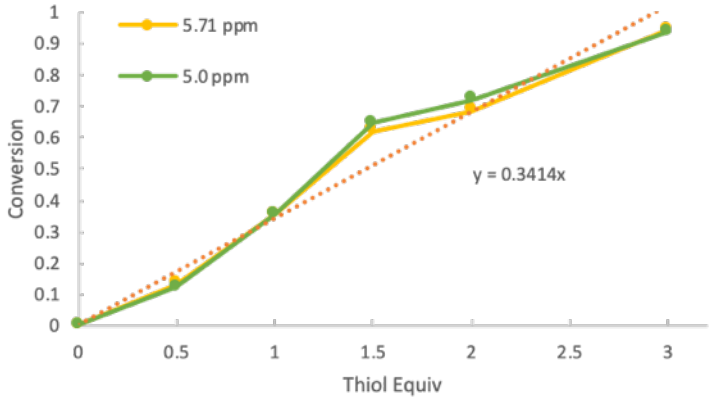
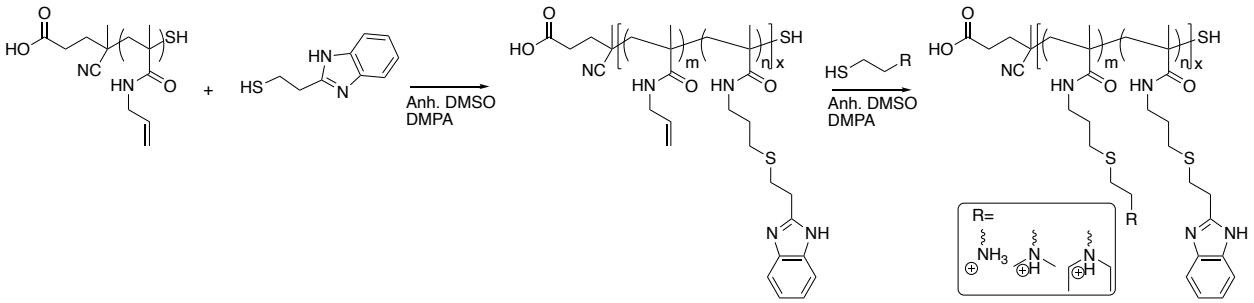


Figure S6. ^{1}H NMR kinetics showing conversion vs thiol equiv. comparing the allyl alkene peaks at 5.7 and 5.0 ppm to an internal standard of phenol (7.1 and 6.7 ppm).

Co-polymer functionalization



Scheme S5. Co-polymer functionalization of pAMAm polymers showing a 2-step, 1- pot thiol-ene click chemistry scheme.

To produce 9 copolymers starting from one pAMAm scaffold the following steps were completed. Each different length of pAMAm (RU: 90, 190, 250, 1 equiv., 1.2 mmol relative to monomeric form, 0.15 g) was dissolved in anhydrous DMSO (0.1 M, 12 mL) and mixed in three varying ratios (0.15 equiv., 0.18 mmol), (0.25 equiv., 0.3 mmol), and (0.5 equiv., 0.6 mmol) of BET. To further the reaction in an open capped scintillation vial, DMPA (0.05 equiv.) was added to the solution mixtures and stirred with a stir bar for 1 h at room temperature in a UV (~365 nm) gel nail curing box. The resulting partially BET-functionalized polymers were each split into 3 vials (4 mL each) and reacted first with one of three thiol-amine salts (cysteamine hydrochloride, captamine hydrochloride or diamino ethanethiol hydrochloride (5 equiv., 2.0 mmol)) and then reacted in a UV nail curing box with DMPA (0.05 equiv.). The resulting polymer solution was acidified with 0.5 mL HCl (1 M) and purified in a 1 kDa dialysis bag in MilliQ-purified water. The polymer solution was further concentrated via lyophilization to obtain a fluffy off-white powder. Polymers with higher incorporation of BET were fluffier. The mass of each polymer obtained was between 70-90 mg. Copolymer molar mass and dispersity measurements were conducted on H_2O SEC-MALLS (0.1 M Na_2SO_4, 1% acetic acid).

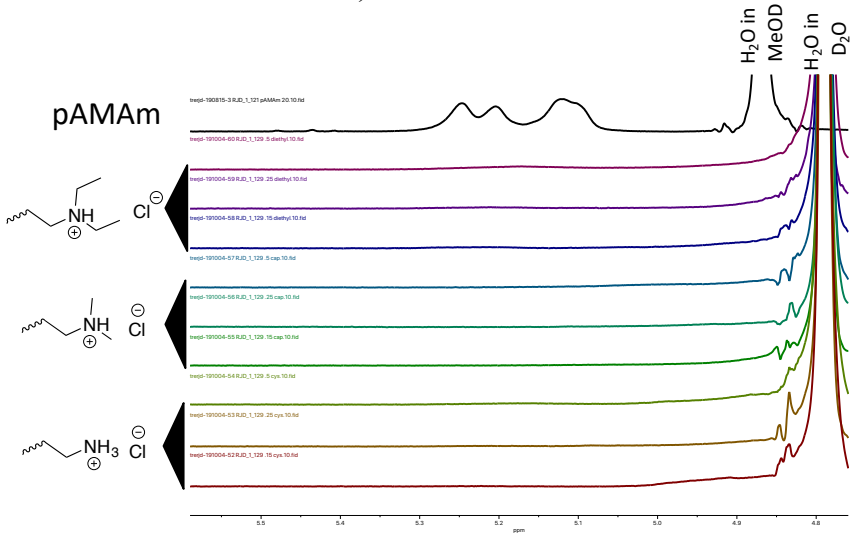


Figure S7. ^{1}H NMR spectra representative stack showing the alkene region (4.8-5.5 ppm) of nine of the functionalized copolymers in the shortest backbone compared to pAMAm material and showing complete depletion of the allyl alkene peak at 5.1 and 5.2 ppm.

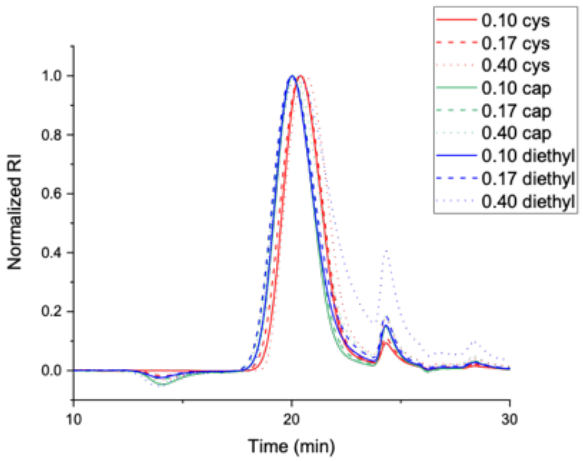


Figure S8. The normalized dRI Aq. SEC-MALS representative stacked trace of nine of the functionalized copolymers after reaction with the shortest pAMAm backbone.

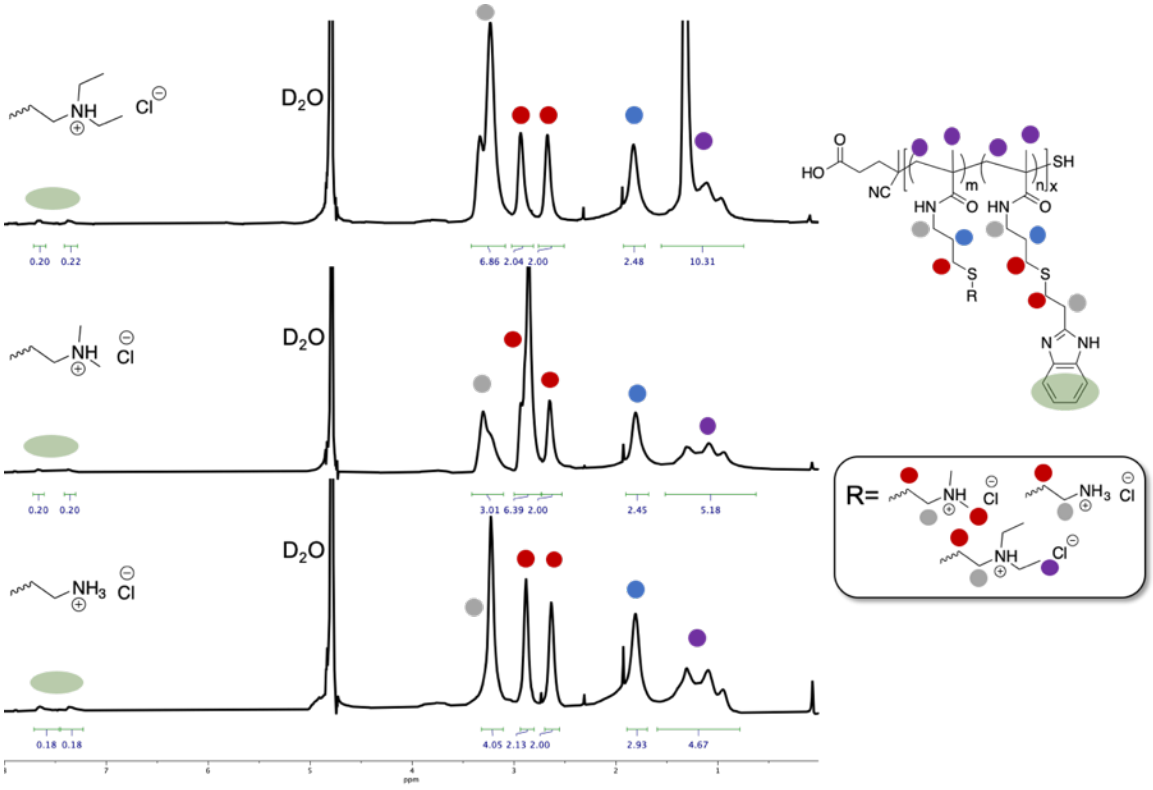


Figure S9. [1]H NMR spectra showing a representative characterization of the pure functionalize copolymers. This shows BET (~10%) and the remaining Cys, Cap, or DiE (~90%) functionalized copolymer.

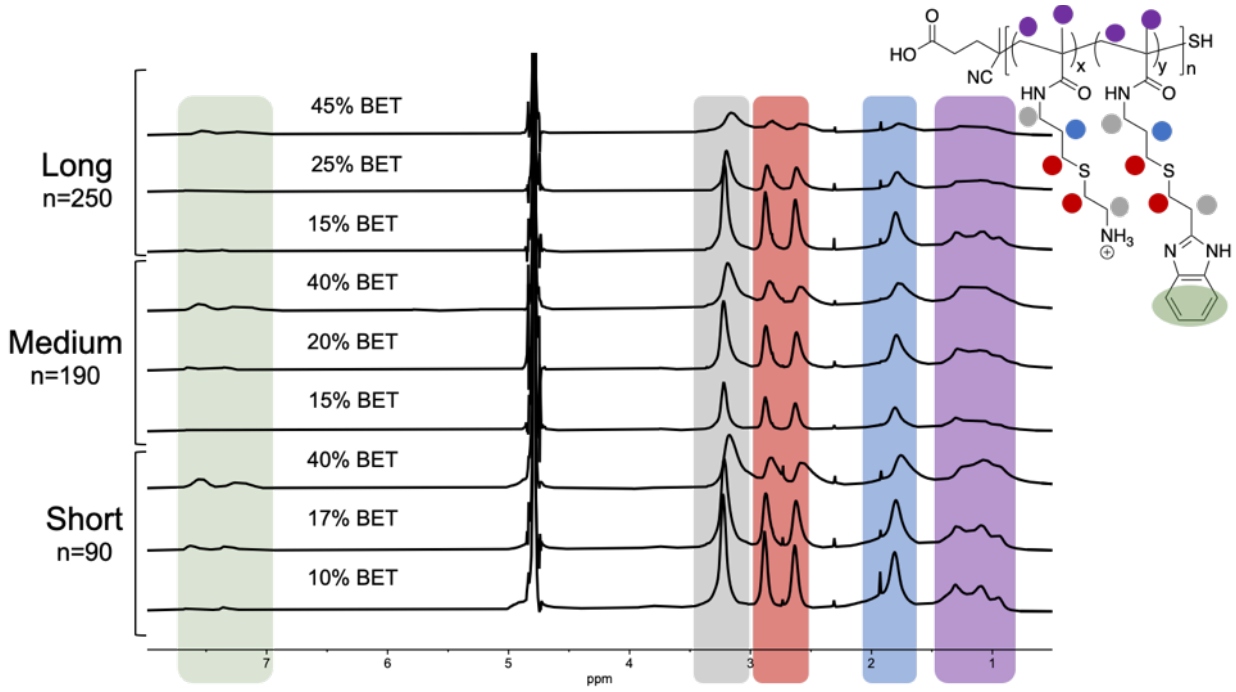


Figure S10. ^{1}H NMR stack spectra showing the cysteamine functionalized copolymers with varying amounts of incorporation of BET with all three lengths of backbone.

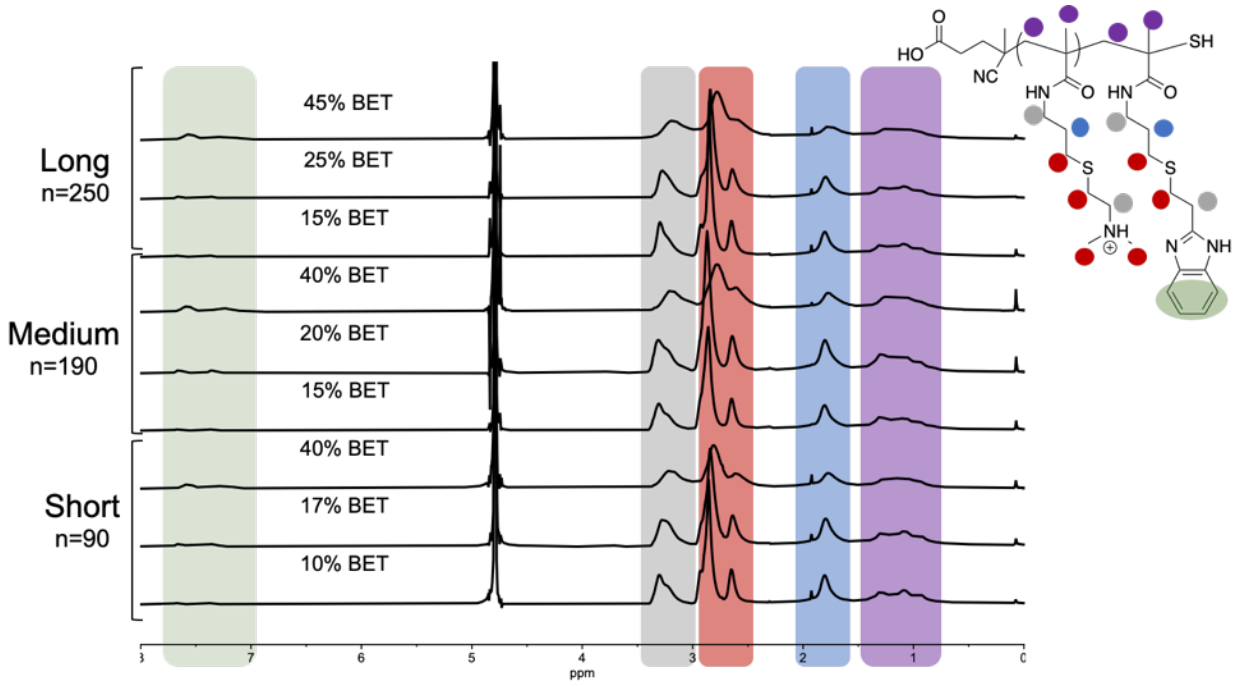


Figure S11. ^{1}H NMR stack spectra showing the captamine functionalized copolymers with varying amounts of incorporation of BET with all three lengths of backbone.

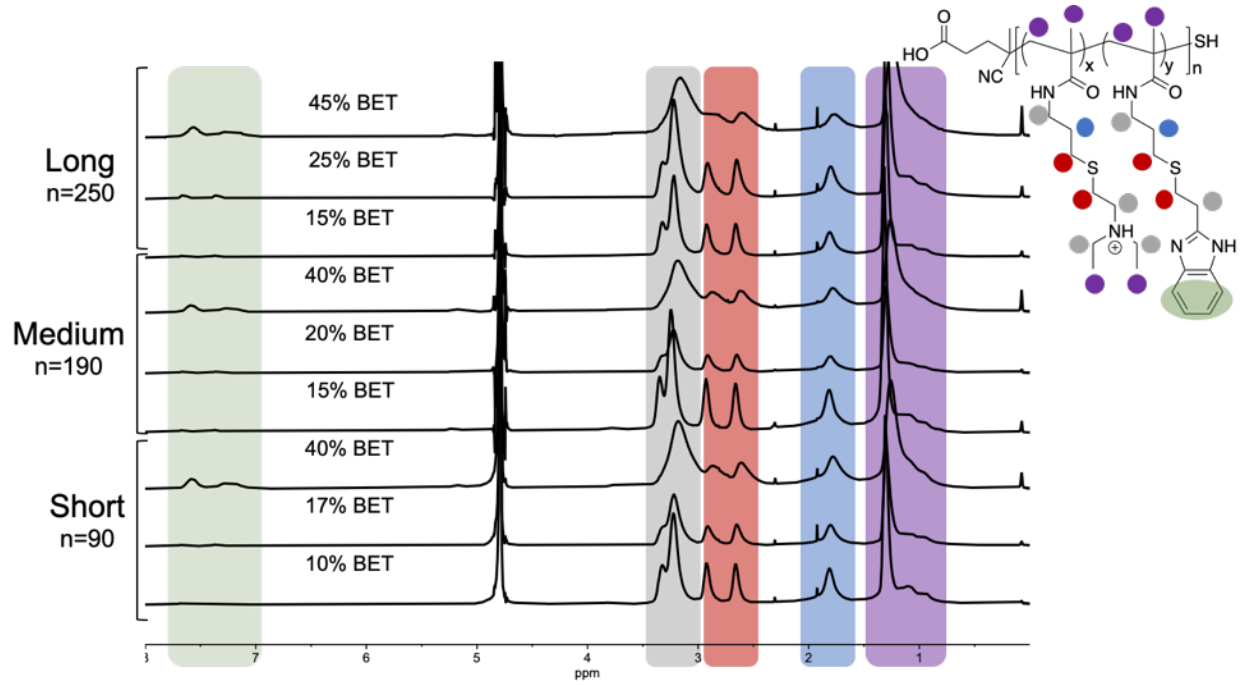


Figure S12. ^{1}H NMR stack spectra showing the diethylamino ethanethiol functionalized copolymers with varying amounts of incorporation of BET with all three lengths of backbone. Table of Mw, Mn, D,

Offline batch mode measurement of ∂n/∂c of macromonomer and bottlebrush

A stock solution of polymer was made in H_2O (0.1 M Na_2SO_4, 1% acetic acid). Five dilutions were made in H_2O (0.1 M Na_2SO_4, 1% acetic acid) to achieve an order of magnitude difference with a concentration range of 0.2-2.0 mg/mL for Sh_Cys_0, Sh_Cap_0, Sh_DiE_0, and Sh_Cys_40. Samples were injected at a flow rate of 0.1 mL/min into a Wyatt Optilab T-rEX refractive index detector (25 °C, λ_0 = 660 nm). Refractive indices were measured at each concentration and the ∂n/∂c was determined from the slope. The four ∂n/∂c measurements (Sh_Cys_0 = 0.1633, Sh_Cap_0 = 0.1795, Sh_DiE_0 = 0.1898, and Sh_Cys_40 = 0.1998) were used to calculate ∂n/∂c's for the remaining polymers based on an equation from Striegel *et al.*[2]

$$\left(\frac{\partial n}{\partial c}\right)_{AB} = \mathrm{w}_A\left(\frac{\partial n}{\partial c}\right)_A + \mathrm{w}_B\left(\frac{\partial n}{\partial c}\right)_B$$

Equation S1. Equation adapted from Striegel *et al.*[2] for calculating dn/dc values for AB copolymers

pK_a titrations of the four thiol small molecules

Each small-molecule thiol was titrated to identify the p*K*a of the amine. The small-molecule thiol of BET, cysteamine hydrochloride, 2-(dimethylamino) ethanethiol hydrochloride, and 2-(diethylamino) ethanethiol hydrochloride were diluted to a 1 mg/mL solution in 30 mM HCl. The samples were further titrated with 75 mM NaOH in an auto-titrator, Orion star T901 pH titrator (Thermo Fisher, Waltham, MA).

BET: 5.90, DiE: 7.68, Cap: 7.74, Cys: 8.10

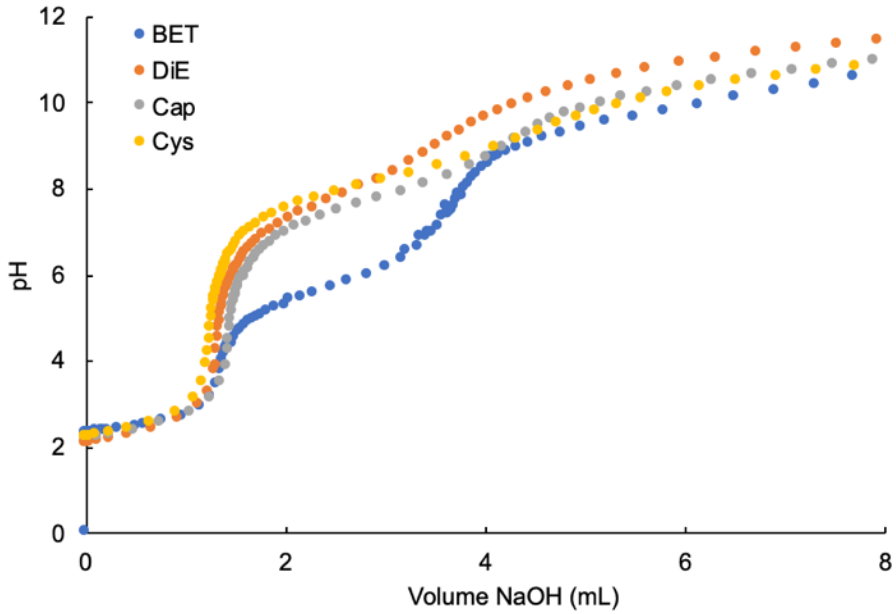


Figure S13. Stacked pK_a titration traces for the four small molecule thiols.

$$Polymer\ pK_a = \frac{(RU)(pK_aBET)(mol\ BET) + (RU)(mol\ cation)(pK_a\ cation)}{RU}$$

Equation S2. Equation to calculate the molar average of pK_a for the polymer systems.

clogP values

The Molinspiration Property Calculation Service, Molinspiration Cheminformatics website (https://www.molinspiration.com/cgi-bin/properties), was utilized to calculate clogP values.

BET: 2.40, DiE: -1.74, Cap: -2.50, Cys: -2.70

$$Polymer\ clogP = \frac{(RU)(mol\ BET)(clogP\ BET) + (RU)(mol\ cation)(clogP\ cation)}{RU}$$

Equation S3. Equation to calculate the molar average of clogP for the polymer systems.

Table S1. Full characterization of entire polymeric library.

Scaffold RU	%BET	%Cation	Cation	clogP	pK_a	M_n	M_w	$Đ$
90 (Sh)	0	100	Cys	-2.70	8.10	13	16	1.21
			Cap	-2.50	7.74	13	16	1.19
			DiE	-1.74	7.68	15	18	1.20
	10	90	Cys	-2.23	7.88	22	26	1.21
			Cap	-2.05	7.56	22	27	1.25
			DiE	-1.36	7.50	22	27	1.23
	17	83	Cys	-1.89	7.73	25	33	1.29
			Cap	-1.73	7.43	24	33	1.33
			DiE	-1.10	7.38	25	32	1.29
	40	60	Cys	-0.80	7.22	31	37	1.22
			Cap	-0.68	7.00	30	40	1.33
			DiE	-0.23	6.97	32	40	1.23
190 (Md)	0	100	Cys	-2.70	8.10	30	42	1.41
			Cap	-2.50	7.74	43	44	1.30
			DiE	-1.74	7.68	35	51	1.45
	15	85	Cys	-1.99	7.77	36	57	1.51
			Cap	-1.82	7.46	30	46	1.50
			DiE	-1.17	7.41	31	46	1.47
	20	80	Cys	-1.75	7.66	43	67	1.55
			Cap	-1.59	7.37	34	53	1.59
			DiE	-0.98	7.32	35	54	1.53
	40	60	Cys	-0.80	7.22	58	78	1.33
			Cap	-0.68	7.00	46	75	1.62
			DiE	-0.23	6.97	56	78	1.41
250 (Lg)	0	100	Cys	-2.70	8.10	52	73	1.40
			Cap	-2.50	7.74	56	71	1.27
			DiE	-1.74	7.68	53	68	1.28
	15	85	Cys	-1.99	7.77	69	114	1.65
			Cap	-1.82	7.46	79	169	2.14
			DiE	-1.17	7.41	73	121	1.66
	25	75	Cys	-1.52	7.55	78	139	1.79
			Cap	-1.37	7.28	84	191	2.27
			DiE	-0.80	7.24	83	149	1.80
	45	55	Cys	-0.57	7.11	195	263	1.35
			Cap	-0.46	6.91	187	382	2.05
			DiE	-0.04	6.88	238	304	1.28

Biological and physical characterization of polyplexes

Cell culture procedures

The HEK293T cell line was cultured in high glucose DMEM with added 10% HI-FBS and 1% antibiotic/antimicrobic. The incubator was set to 37 °C with 5% CO_2 and under a humidified atmosphere. Cell confluency was monitored, and cells were passaged as needed. Cells were plated in a 48-well plate format at a density of 50,000 cells/mL with 0.5 mL of suspension added per well.

General pDNA and RNP polymer transfection protocol

HEK293T cells were plated in polystyrene 48-well plates at a density of 50,000 cells/mL in DMEM (10% HI-FBS). After 24 h, polyplexes with pDNA were prepared in H_2O by adding 80 µL polymer to 80 µL pDNA (0.02 µg/mL) at molar ratios of polymer to get N/P ratios. Polyplexes with RNP were prepared in PBS by adding 80 µL polymer to 80 µL RNP (gRNA (0.02 µg/mL) and spCas9 protein (0.1 µg/mL) complex) at molar ratios of polymer to get N/P ratios. Polyplexes were allowed to form at room temperature for 40 min. Opti-MEM in a 2:1 ratio (320 µL) was added to the polyplexes immediately before addition to cells.

Media was carefully aspirated from the well plate before addition of the polyplex samples. Each polyplex was split into triplicate adding 150 µL to each well. Well plates remained on the bench top at room temperature for 40 min before placing into the 37 °C incubator (5% CO_2). 4 h after initial transfection, 1 mL of DMEM (10% HI-FBS) was added to each well. Media was further aspirated 24 hours after initial transfection, and fresh DMEM (10% HI-FBS, 1 mL) was added to each well. 48 hours after the initial transfection, the cells were analyzed using CCK-8 and flow cytometry, details are listed below.

General flow workup for quantification of expression

Flow cytometry was used to measure and quantify the transfection efficiency of the full polymer library to deliver pDNA encoding for GFP to HEK293T cells. To quantify the percentage of GFP or mCherry positive cells using flow cytometry, cells were harvested 48 h after the initial transfection. To harvest the cells, the media was aspirated and the HEK293T cells were trypsinized (150 µL). The cells were then added to a V-shaped 96-deepwell plate and centrifuged at 4 °C. The supernatant was aspirated away, and cell pellet was resuspended with Calcein Violet live-cell stain in PBS with 1% FBS. The cells were incubated with the cell stain for 30 min on ice without light before measuring via flow cytometry for GFP or mCherry. 405 nm (calcein violet) and 488 nm (GFP) lasers were used on the flow cytometer (ZE5, Biorad, Inc., CA, USA). At least 5,000 events were collected for every treatment in triplicate.

General CCK-8 viability assay

After transfection, cell viability was measured using CCK-8. Transfection procedures were carried out as written in *in vitro* cell transfection using polyplex section. 48 hours post initial transfection, the media was aspirated. Fluorobrite (0.5 mL) containing 10% FBS was mixed with CCK-8 dye (40 µL) and added to each well to incubate for 2 h. After incubation, 150 µL of the supernatant was removed and placed in a 96-well plate and absorbance was measured at 450 nm using a Synergy H1 Hybrid Reader (BioTek, Winooski, VT). Untreated cells were normalized to 1.0 cell survival.

Dynamic light scattering (DLS) with pDNA

The polyplexes were measured by DLS to better understand the hydrodynamic radius (R_h). All PBS (pH= 7.4) were pre-filtered through a 0.2 µm GHP syringe filter and prepared to run in a high-throughput DLS DynaPro Plate Reader III (Greiner Bio One GmbH, SensoPlate, 655892, Wyatt Technology, Santa Barbara, CA). For high-throughput DLS, samples were transferred into a glass-bottomed 96-well DLS plate. The well plate was placed in the DynaPro Plate Reader III and equilibrated at 25 °C. Polyplexes were prepared in PBS by adding 100 µL polymer to 100 µL pDNA (0.02 µg/mL) at a molar ratio of polymer to achieve N/P ratios of 10:1 and 20:1. After letting the polyplex form for 40 min in the well, samples of interest were analyzed using automated measurements. For each measurement, five acquisitions were recorded with an acquisition time of 5 seconds each. A naked plasmid control was measured and had a hydrodynamic radius of 99.3 ± 1.9 nm.

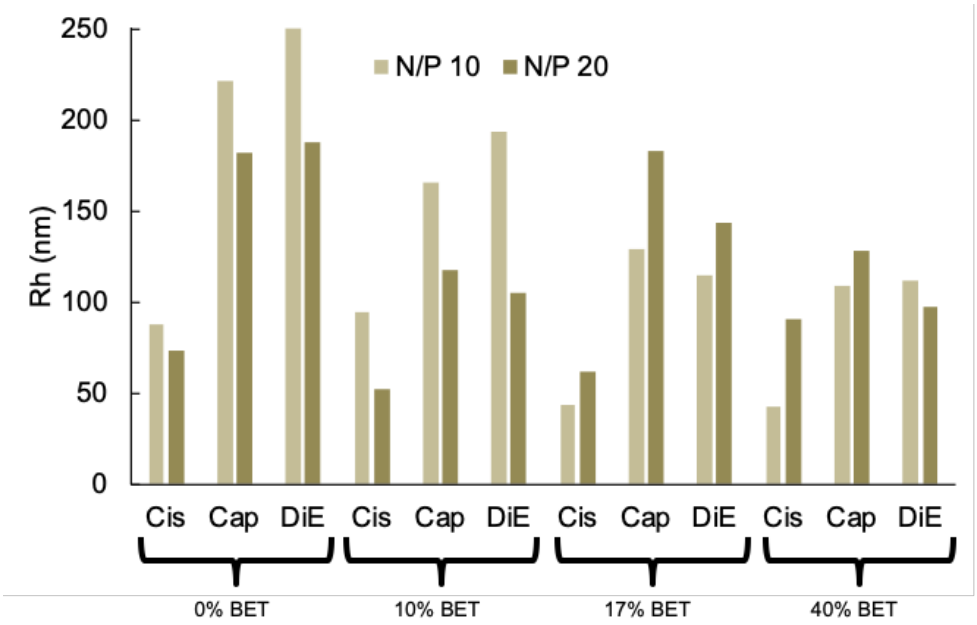


Figure S14. DLS data showing the intensity average hydrodynamic radius of the polyplexes formed in PBS with the short length backbone functionalized polymers at an N/P ratio of 10 and 20.

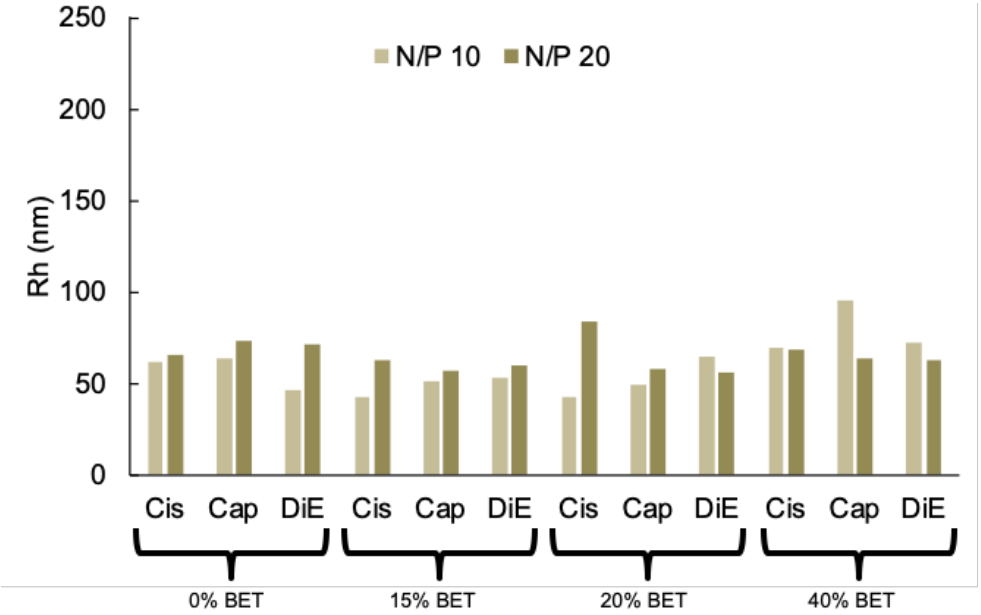


Figure S15. DLS data showing the intensity average hydrodynamic radius of the polyplexes formed in PBS with the medium length backbone functionalized polymers at an N/P ratio of 10 and 20.

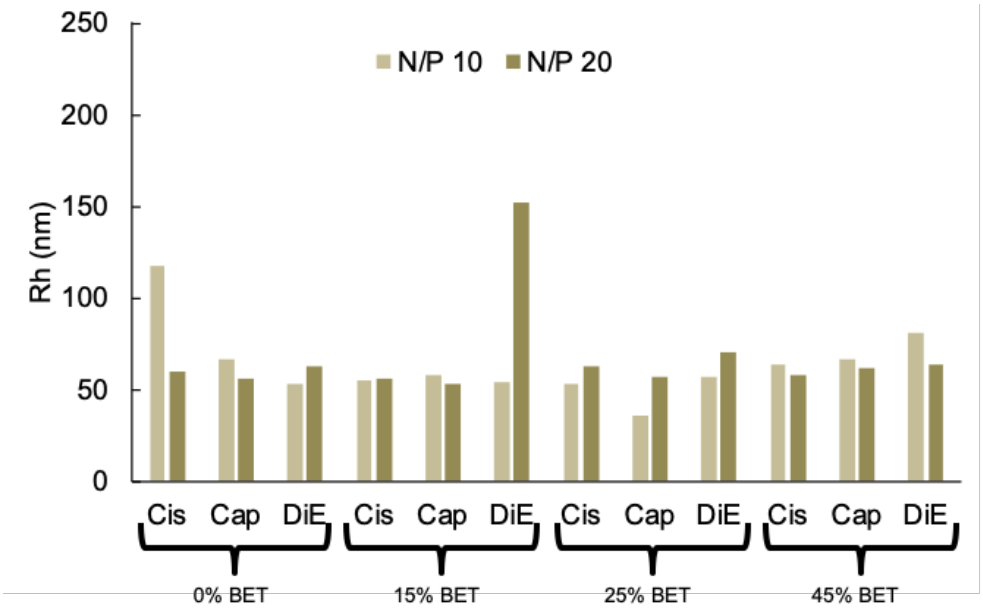


Figure S16. DLS data showing the intensity average hydrodynamic radius of the polyplexes formed in PBS with the long length backbone functionalized polymers at an N/P ratio of 10 and 20.

Dynamic light scattering (DLS) with RNP

The hydrodynamic radii of polyplexes from RNP and backbone functionalized polymers in PBS (pH = 7.4) were measured by DLS with a Zetasizer Nano Series (Malvern analytical, Westborough, MA). PBS was pre-filtered through a 0.2 µm syringe filter. To prepare RNP, 30 µL of gRNA (0.04 mg/mL) was added to 30 µL of spCas9 (0.2 mg/mL) and allowed to form complexes for 20 minutes. To form polyplexes, 60 µL of the polymer solution at a molar ratio of polymer to achieve an N/P ratio of 5 was added. Polyplexes were allowed to form for 45 minutes and diluted with 240 µL of PBS. Samples were placed in disposable microcuvettes, and 3 measurements were taken using 173° backscatter.

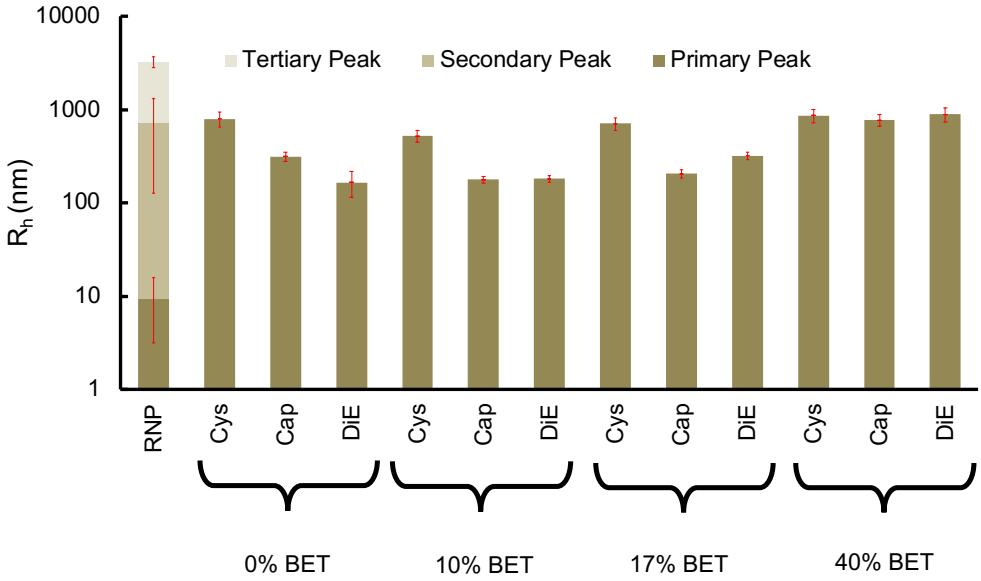


Figure S17. DLS data showing the intensity average hydrodynamic radius of the polyplexes formed in PBS with the short length backbone functionalized polymers at an N/P ratio of 5.

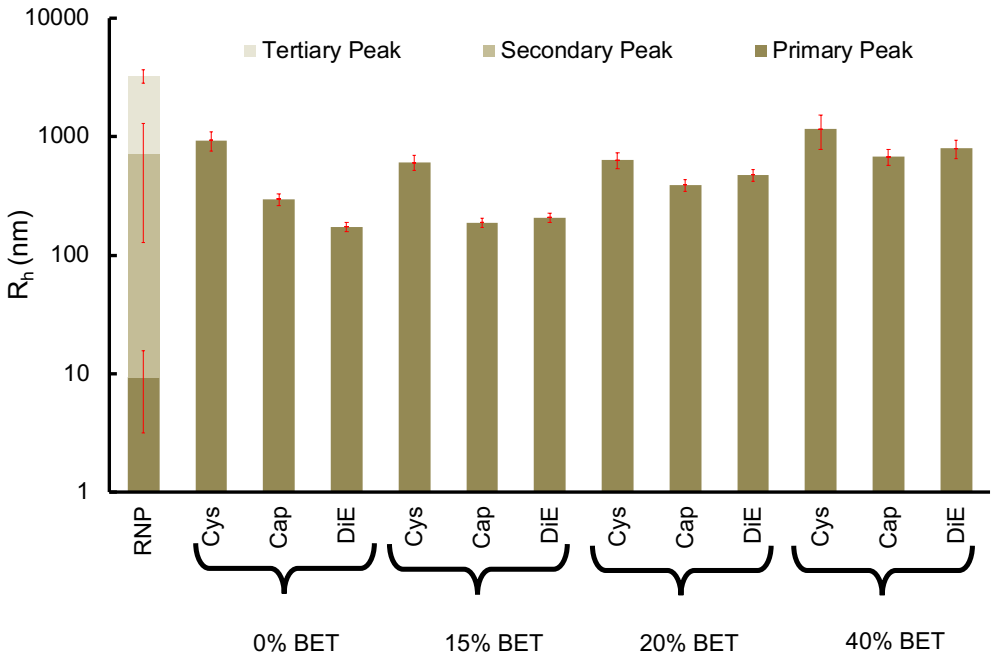


Figure S18. DLS data showing the intensity average hydrodynamic radius of the polyplexes formed in PBS with the medium length backbone functionalized polymers at an N/P ratio of

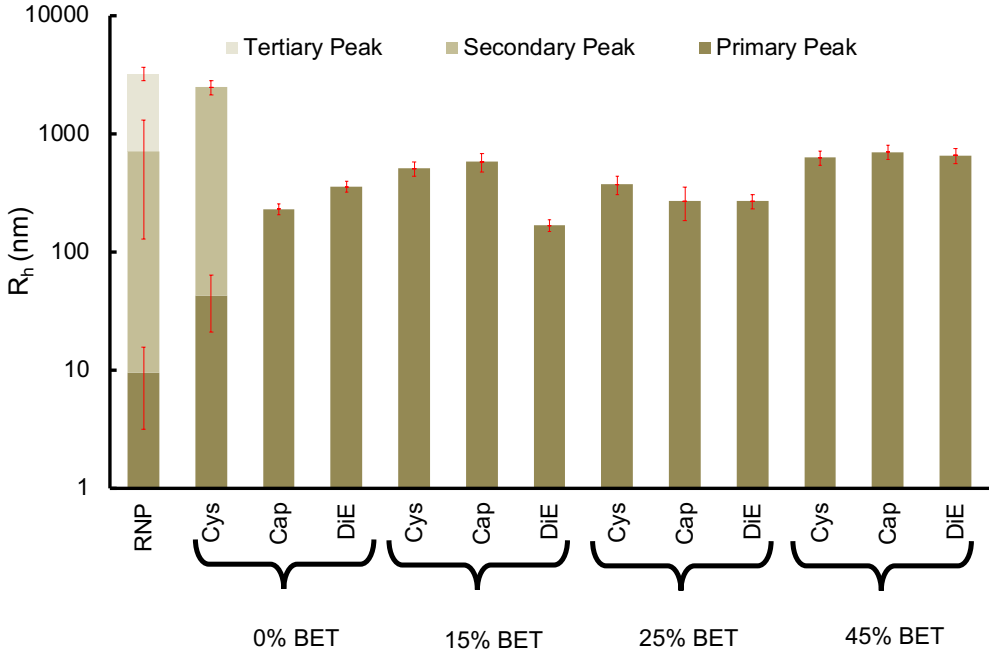


Figure S19. DLS data showing the intensity average hydrodynamic radius of the polyplexes formed in PBS with the long length backbone functionalized polymers at an N/P ratio of 5.

Dye exclusion with pDNA

The small molecule PicoGreen dye was incubated for 15 min with pDNA (0.02 mg/mL) and was diluted to a final concentration with a ratio of 1:200 in H_2O. Polyplexes were prepared in H_2O by adding 170 µL polymer to 170 µL PicoGreen-pDNA (0.02 mg/mL) at various molar ratios of polymer to get N/P ratios of 10:1 and 20:1. Polyplexes were allowed to form at room temperature for 40 min and then split into triplicate by separating the polyplex solution into three wells of 100 µL each in a 96 well plate. Fluorescence of PicoGreen was measured (excitation: 485 nm, emission: 528 nm) using a Synergy H1 Hybrid Reader (BioTek, Winooski, VT). Since PicoGreen only fluoresces when intercalated with the pDNA, the lack of fluorescence signifies exclusion and thus, polymer binding to the pDNA. Samples were normalized in comparison to pDNA and PicoGreen with no competitive polymer. A blank control of PicoGreen in water was also run and subtracted from each sample to account for autofluorescence/background.

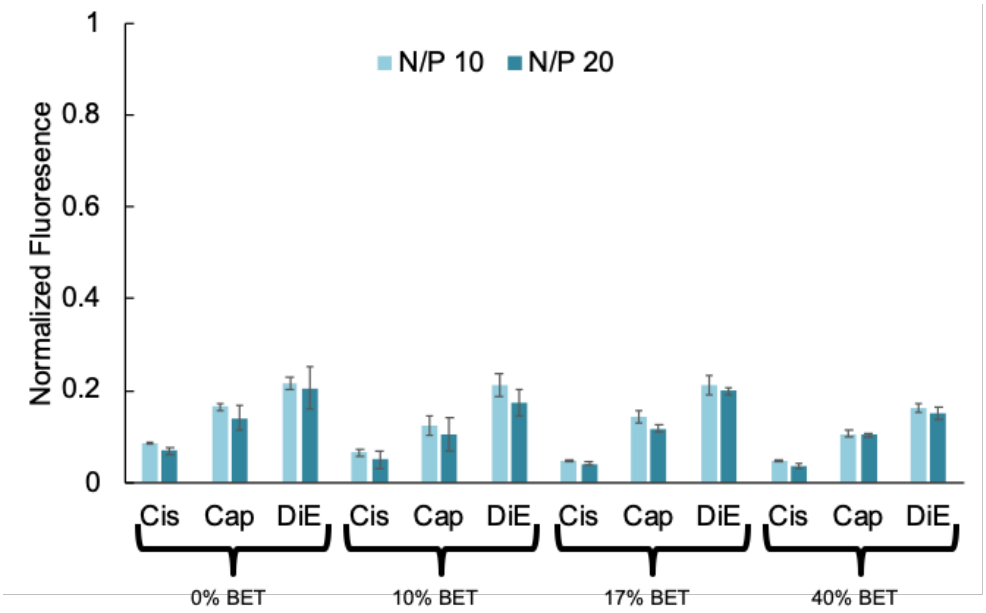


Figure S20. Dye exclusion data showing the amount of fluorescence of PicoGreen after the addition of the short length backbone functionalized polymers at an N/P ratio of 10 and 20.

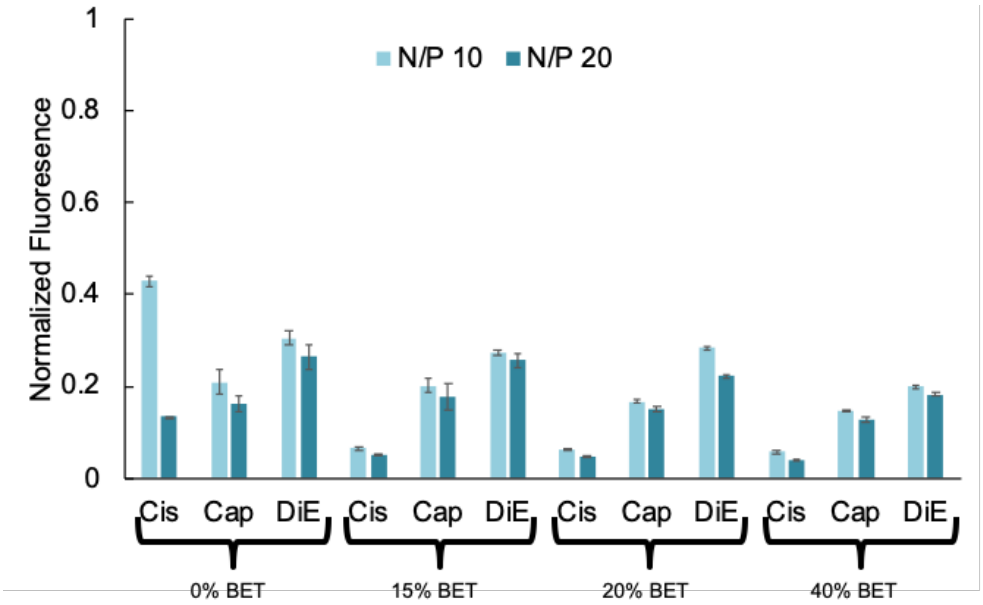


Figure S21. Dye exclusion data showing the amount of fluorescence of PicoGreen after the addition of the medium length backbone functionalized polymers at an N/P ratio of 10 and 20.

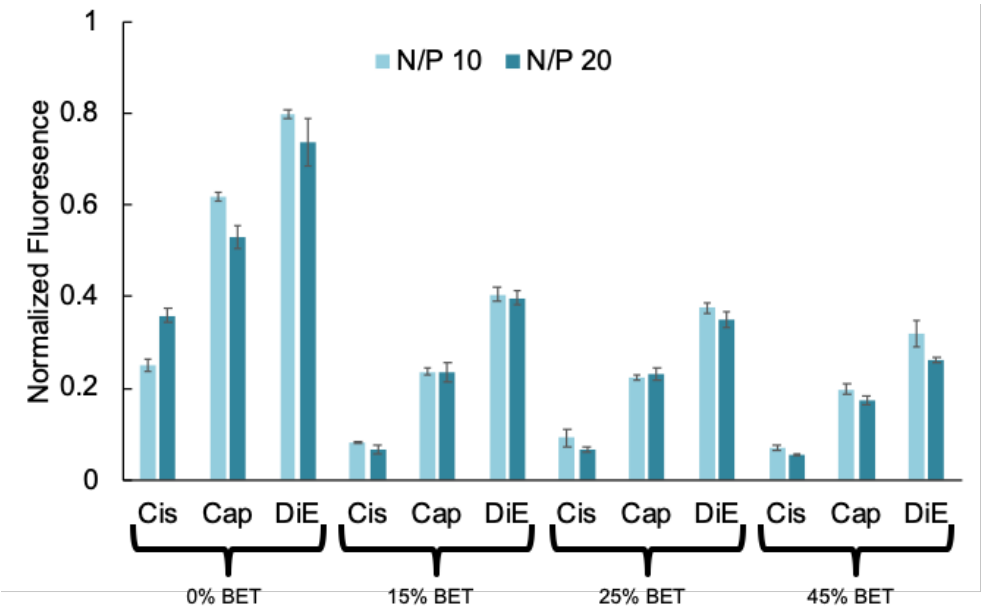


Figure S22. Dye exclusion data showing the amount of fluorescence of PicoGreen after the addition of the long length backbone functionalized polymers at an N/P ratio of 10 and 20.

Dye exclusion with RNP

The small molecule OliGreen dye was incubated for 15 min with RNP (gRNA (0.02 µg/mL) and spCas9 protein (0.1 µg/mL) complex) and was diluted to a final concentration with a ratio of 1:200 in PBS. Polyplexes were prepared in PBS by adding 170 µL polymer to 170 µL OliGreen-RNP at a molar ratio of polymer to get an N/P ratio of 5:1. Polyplexes were allowed to form at room temperature for 40 min and then were split into triplicate by separating the polyplex solution into three wells of 100 µL each in a 96 well plate. Fluorescence of OliGreen was measured (excitation: 485 nm, emission: 528 nm) using a Synergy H1 Hybrid Reader (BioTek, Winooski, VT). Since OliGreen only fluoresces when intercalated with the pDNA, the lack of fluorescence signifies exclusion and thus, polymer binding to the pDNA. Samples were normalized in comparison to RNP and OliGreen with no competitive polymer. A blank control of OliGreen in water was also run and subtracted from each sample to account for autofluorescence/background.

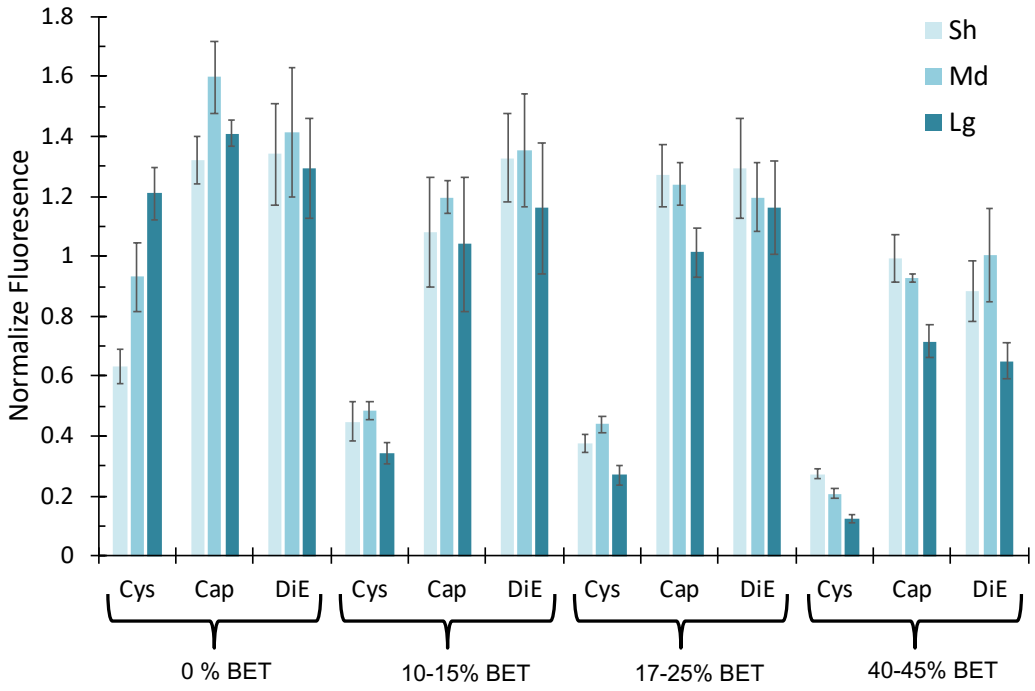


Figure S23. Dye exclusion data showing the amount of fluorescence of OliGreen after the addition of the functionalized polymers at an N/P ratio of 5.

Round 1 pDNA and mCherry transfection

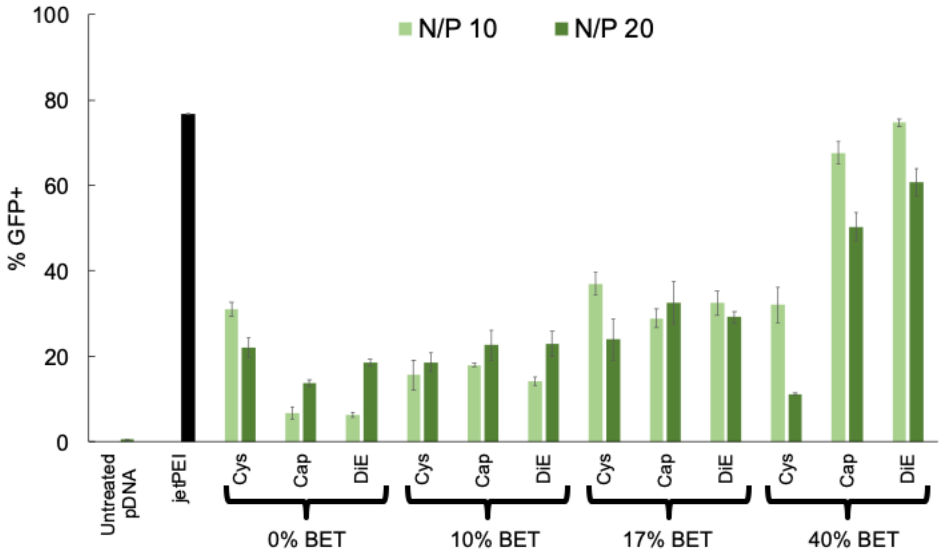


Figure S24. Flow cytometry data showing the output of percent GFP+ cells when the short backbone polymers delivered pDNA into HEK293T cells at an N/P ratio of 10 and 20.

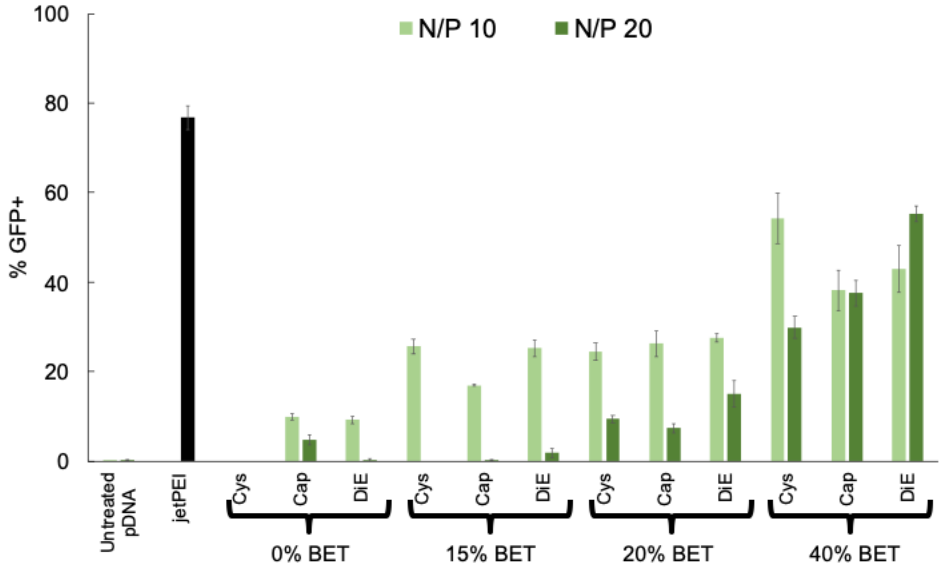


Figure S25. Flow cytometry data showing the output of percent GFP+ cells when the medium backbone polymers delivered pDNA into HEK293T cells at an N/P ratio of 10 and 20.

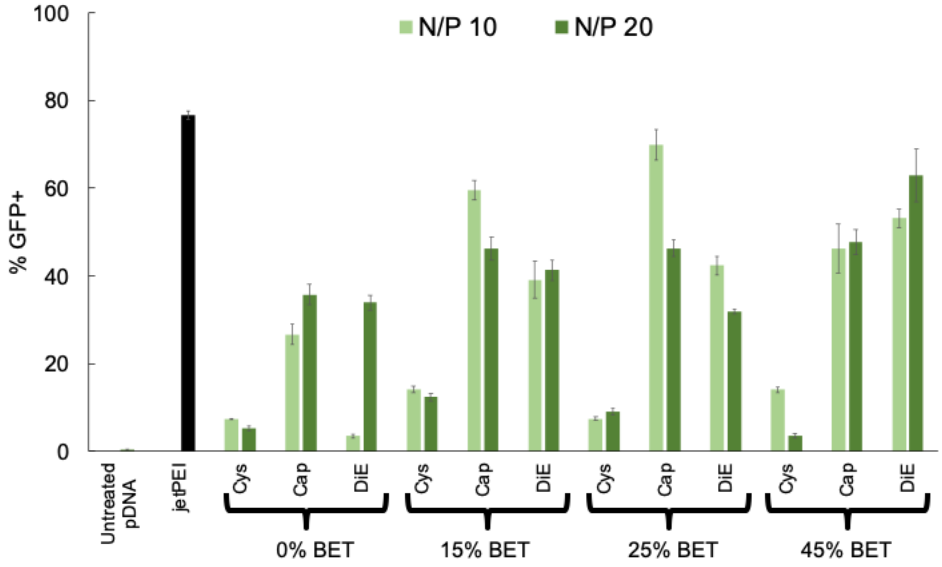


Figure S26. Flow cytometry data showing the output of percent GFP+ cells when the long backbone polymers delivered pDNA into HEK293T cells at an N/P ratio of 10 and 20.

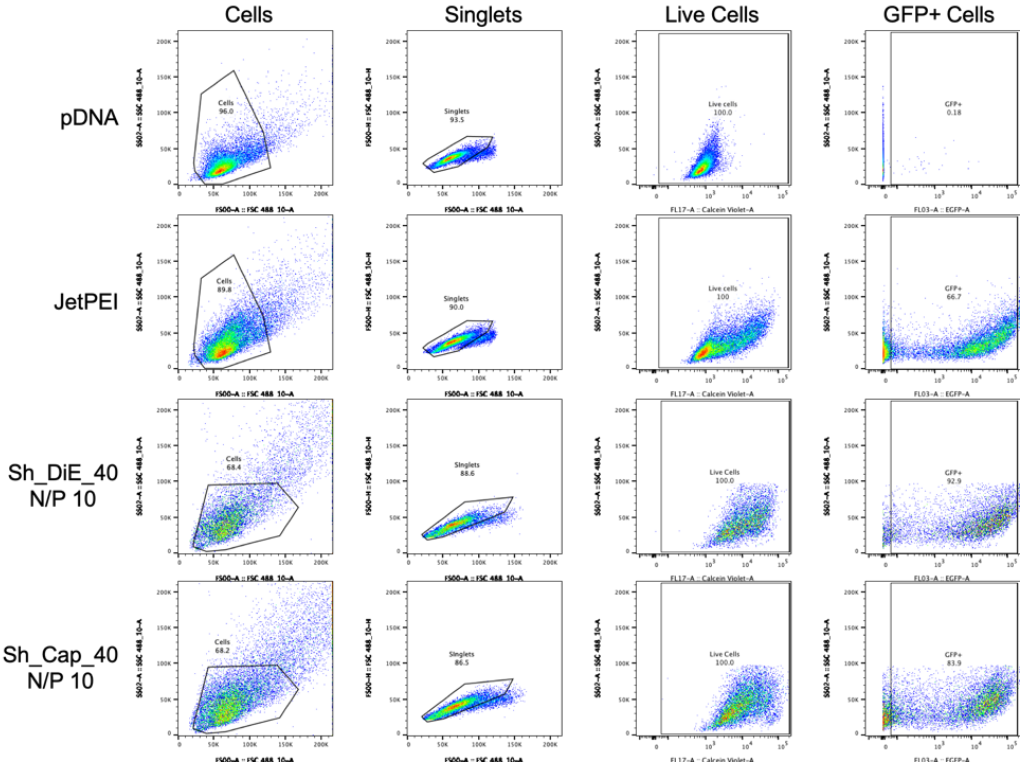


Figure S27. Gating scheme on the flow cytometer used for the identifying the raw data output of percent GFP+ cells. Above shows the gating for Cells, Singlets, Live Cells, and GFP+ cells for four representative samples of pDNA only, JetPEI, Sh_DiE_40, and Sh_Cap_40 before normalization of data.

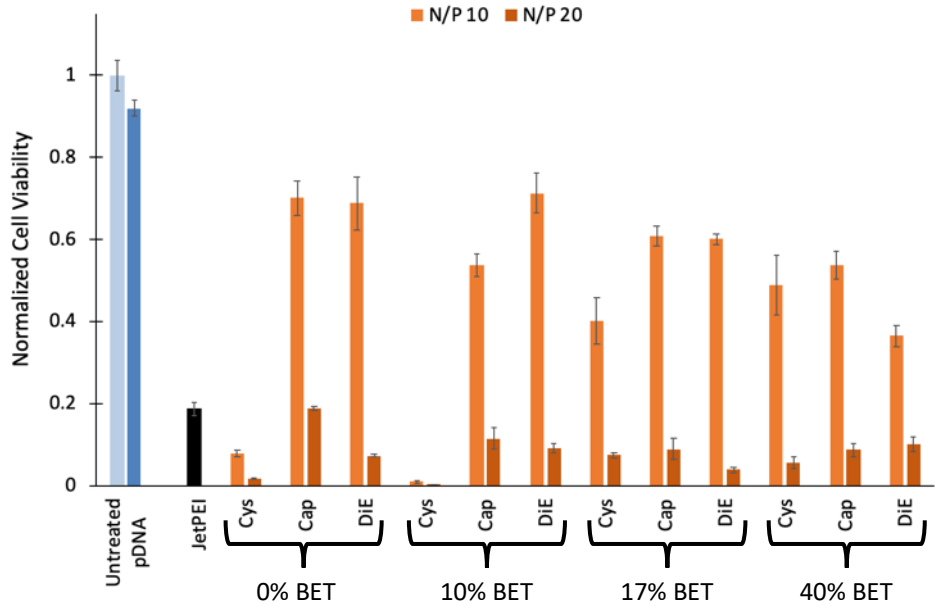


Figure S28. Viability data showing the normalized transmittance in a CCK8 assay when the short backbone polymers delivered RNP into HEK293T cells at an N/P ratio of 10 and 20.

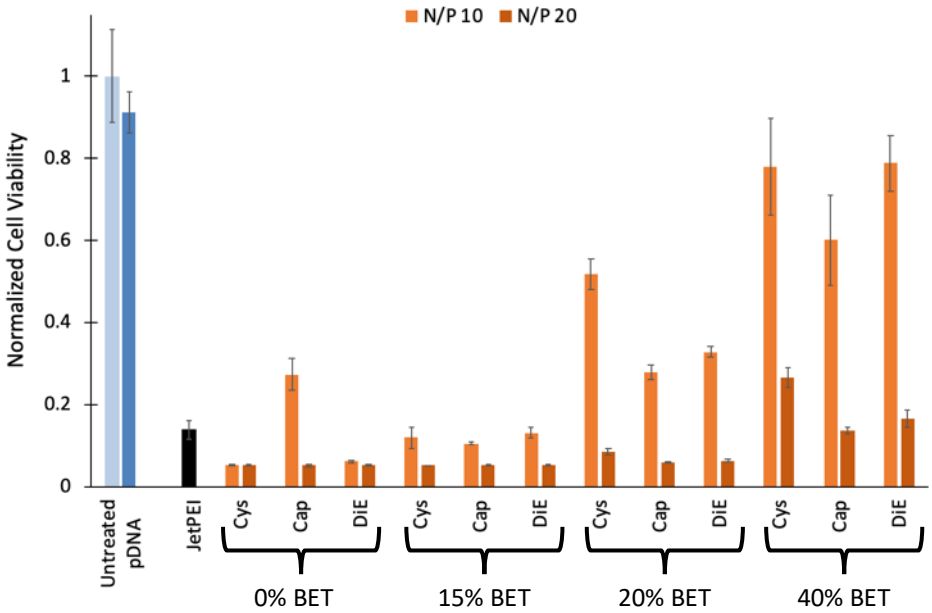


Figure S29. Viability data showing the normalized transmittance in a CCK8 assay when the medium backbone polymers delivered RNP into HEK293T cells at an N/P ratio of 10 and 20.

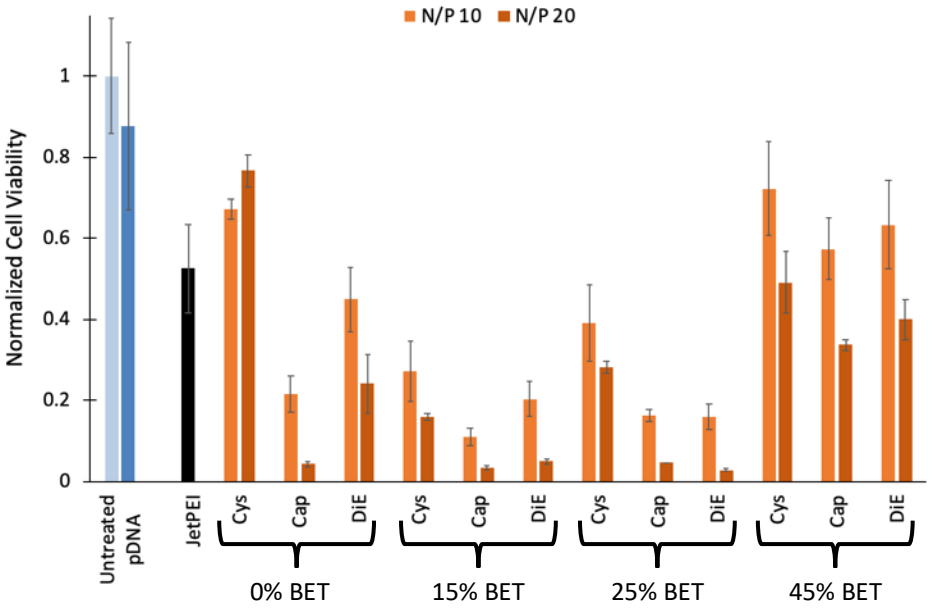


Figure S30. Viability data showing the normalized transmittance in a CCK8 assay when the long backbone polymers delivered RNP into HEK293T cells at an N/P ratio of 10 and 20.

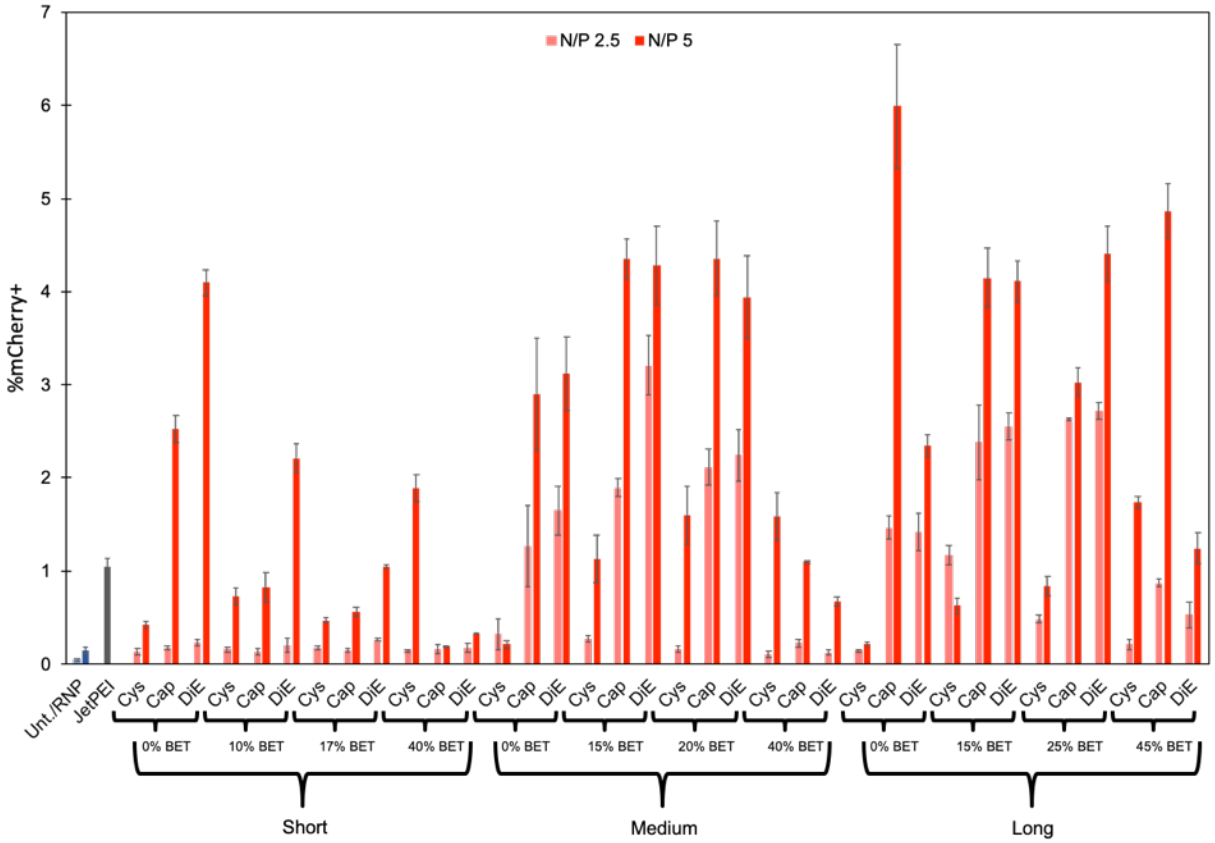


Figure S31. Flow cytometry data showing the output of percent mCherry+ cells with the entire library of polymers delivered RNP into HEK293T cells at an N/P ratio of 2.5 and 5.

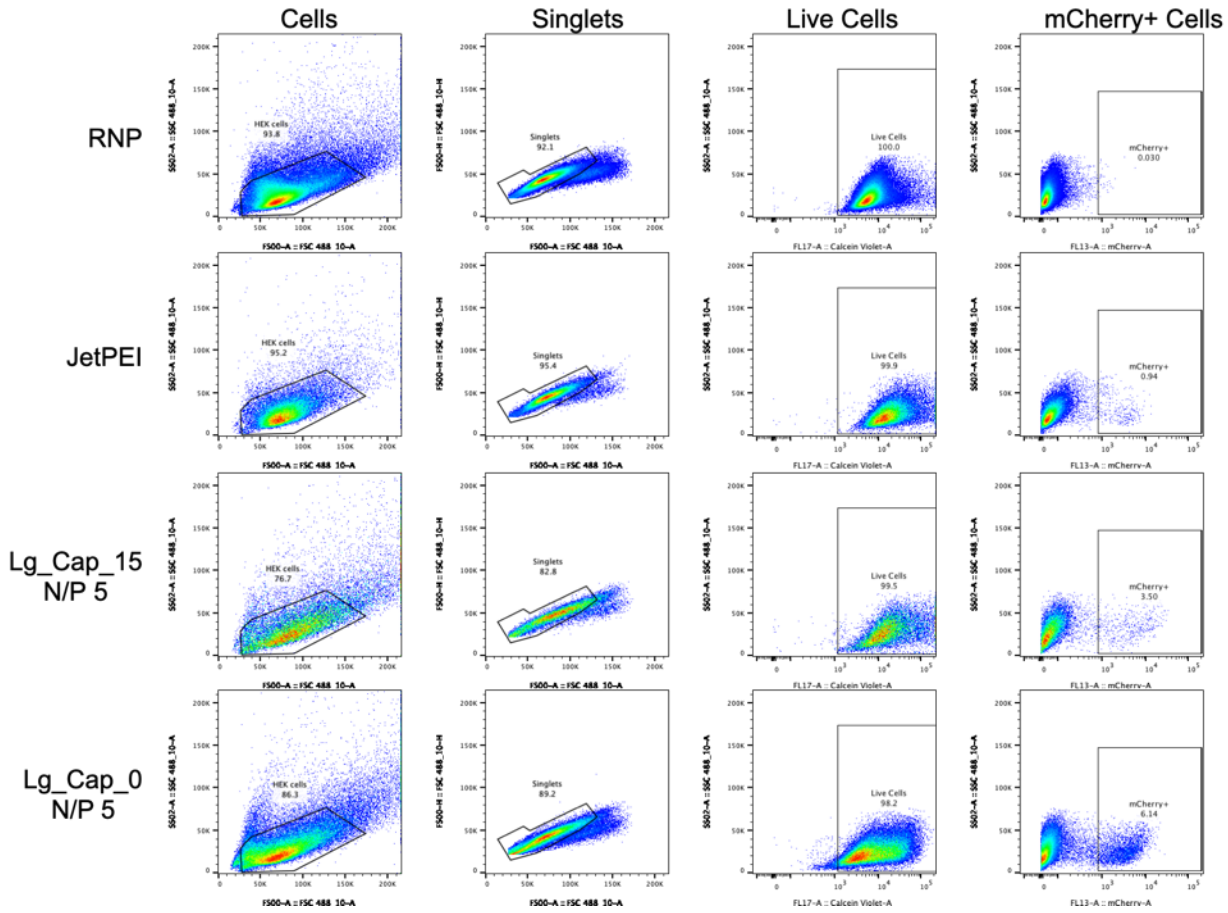


Figure S32. Gating scheme on the flow cytometer used for the identifying the raw data output of percent mCherry+ cells. Above shows the gating for Cells, Singlets, Live Cells, and mCherry+ cells for four representative samples of pDNA only, JetPEI, Lg_Cap_15, and Lg_Cap_0 before normalization of data.

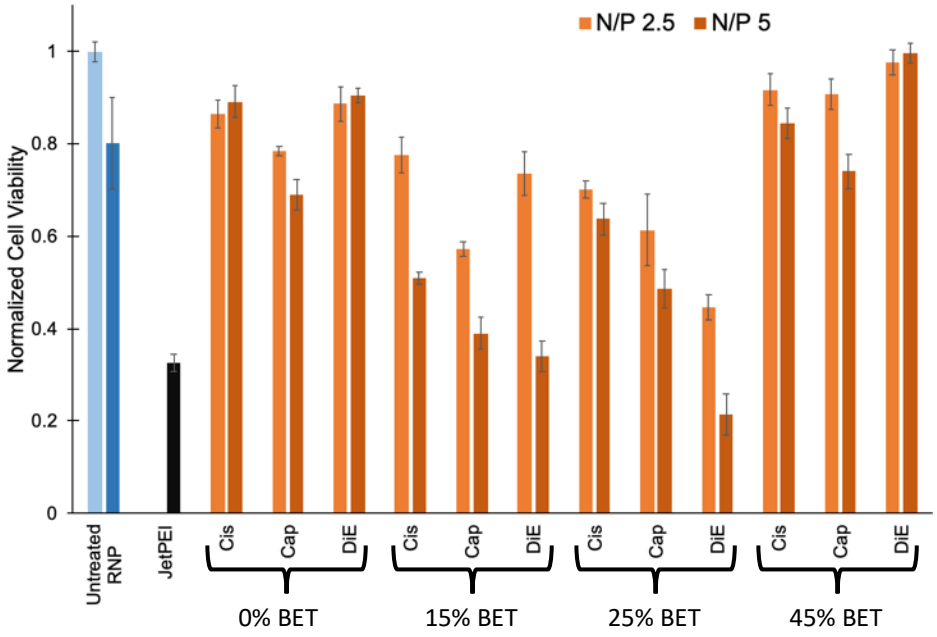


Figure S33. Viability data showing the normalized transmittance in a CCK8 assay when the short backbone polymers delivered RNP into HEK293T cells at an N/P ratio of 2.5 and 5.

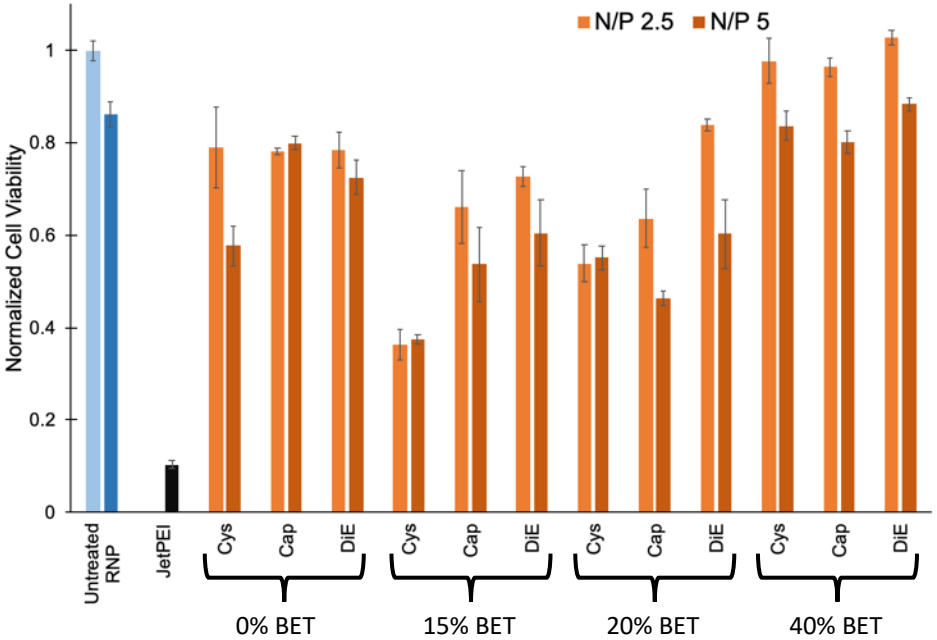


Figure S34. Viability data showing the normalized transmittance in a CCK8 assay when the medium backbone polymers delivered RNP into HEK293T cells at an N/P ratio of 2.5 and 5.

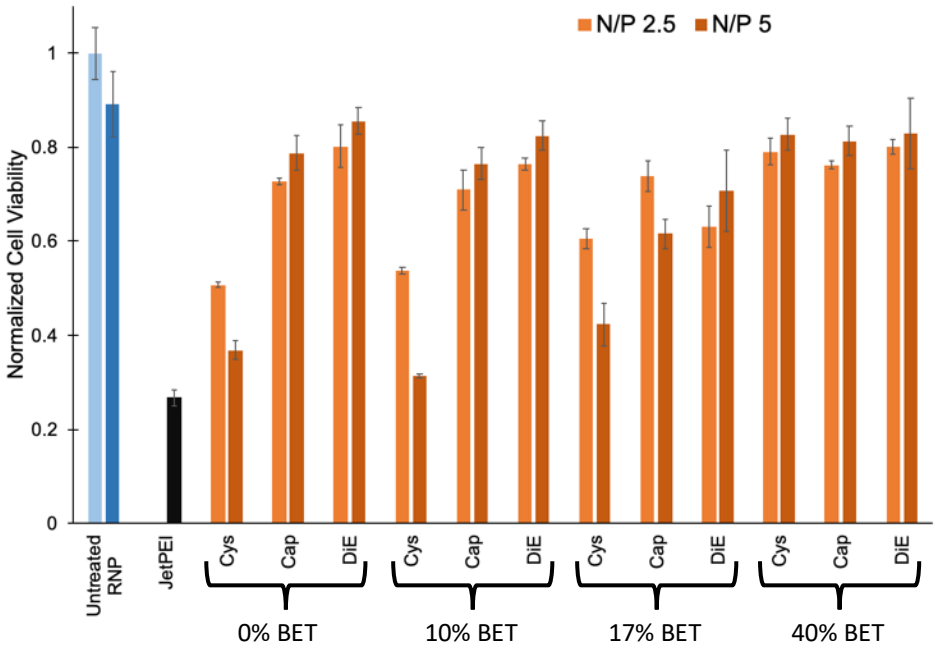


Figure S35. Viability data showing the normalized transmittance in a CCK8 assay when the long backbone polymers delivered RNP into HEK293T cells at an N/P ratio of 2.5 and 5.

Round 2 pDNA and mCherry transfection

Based on predictions of the Bayesian machine learning model, six polymers for each biological payload were chosen for the round two study at two defined N/P ratios. Polyplex formation, transfection, CCK-8 analysis, and flow cytometry workup were all followed as listed above in the general procedures. For pDNA delivery, polymers of Sh_DiE_40, Sh_Cap_40, Lg_Cap_25, Md_Cys_40, Md_DiE_40, and Lg_DiE_45 were tested at N/P ratio of 7.5 and 12.5. For RNP delivery, polymers of Lg_Cap_0, Lg_Cap_45, Md_Cap_15, Md_Cap_20, Md_DiE_15, and Sh_DiE_0 were tested at N/P ratio of 7.5 and 10.

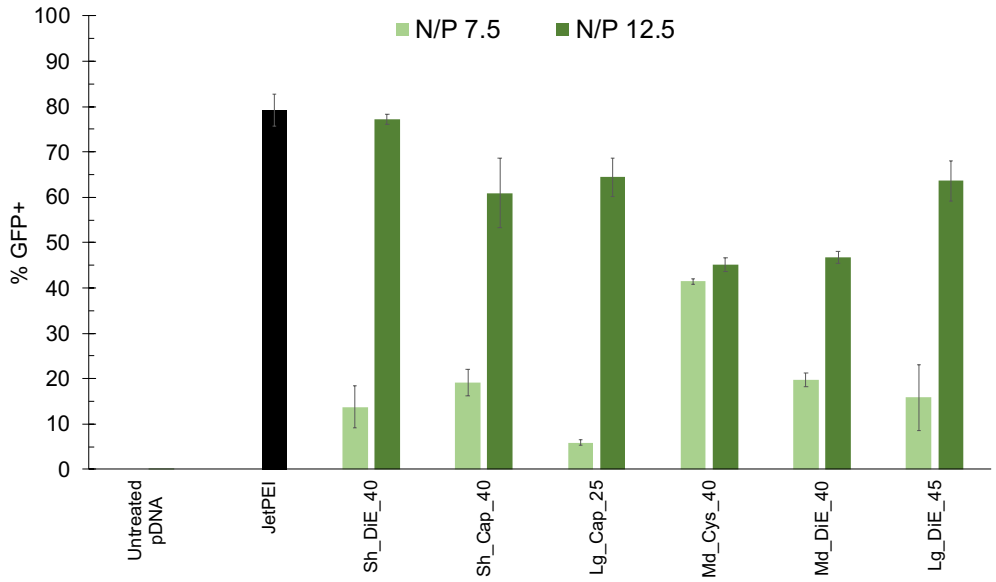


Figure S36. Flow cytometry data showing the output of percent GFP+ cells in our round 2 loop study with various select polymers delivering pDNA into HEK293T cells at an N/P ratio of 7.5 and 12.5.

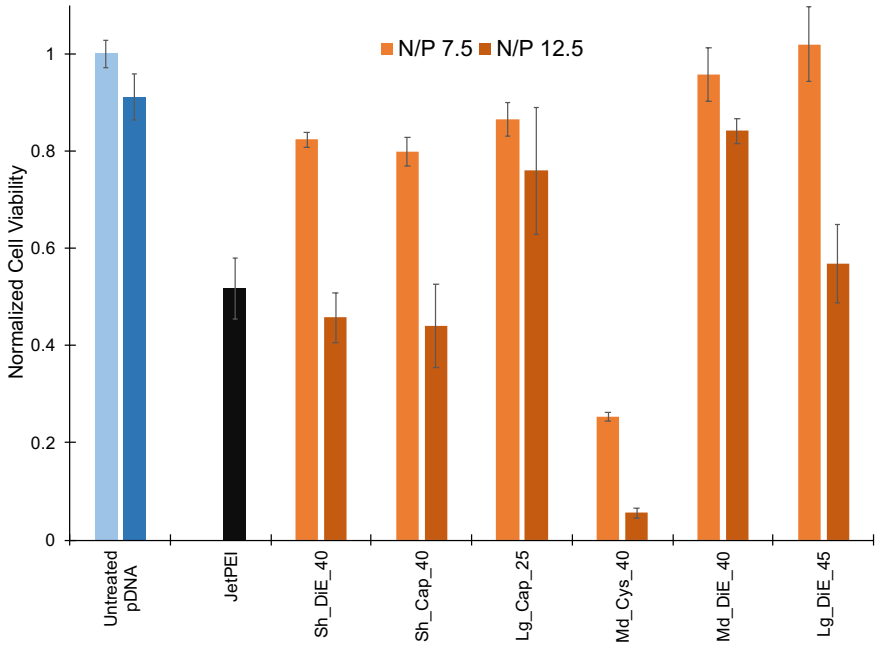


Figure S37. Viability data showing the normalized transmittance in a CCK8 assay in round 2 loop study with select polymers delivering pDNA into HEK293T cells at an N/P ratio of 7.5 and 12.5.

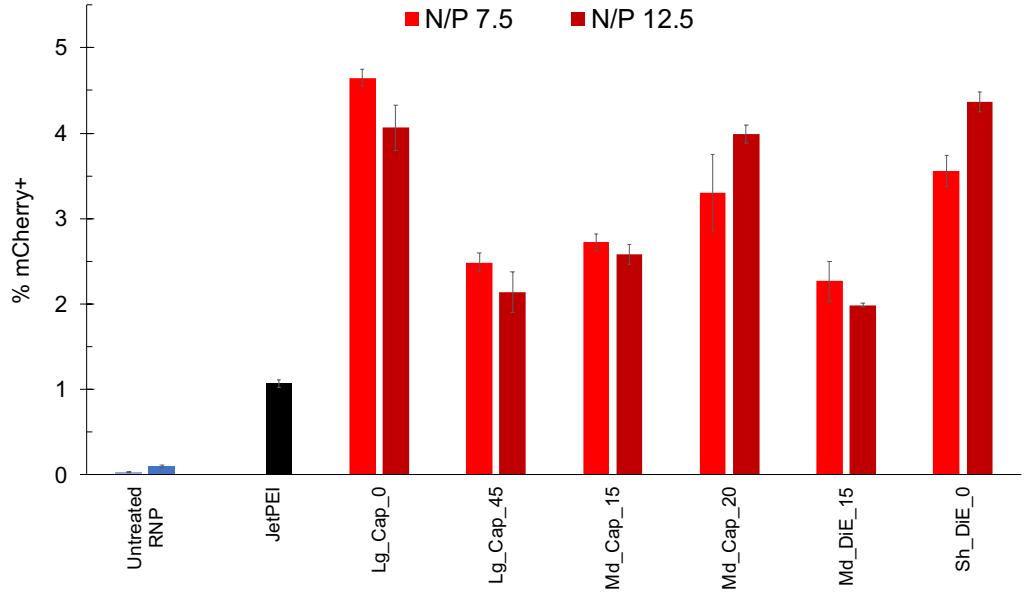


Figure S38. Flow cytometry data showing the output of percent mCherry+ cells in our round 2 loop study with select polymers delivering RNP into HEK293T cells at an N/P ratio of 7.5 and 10.

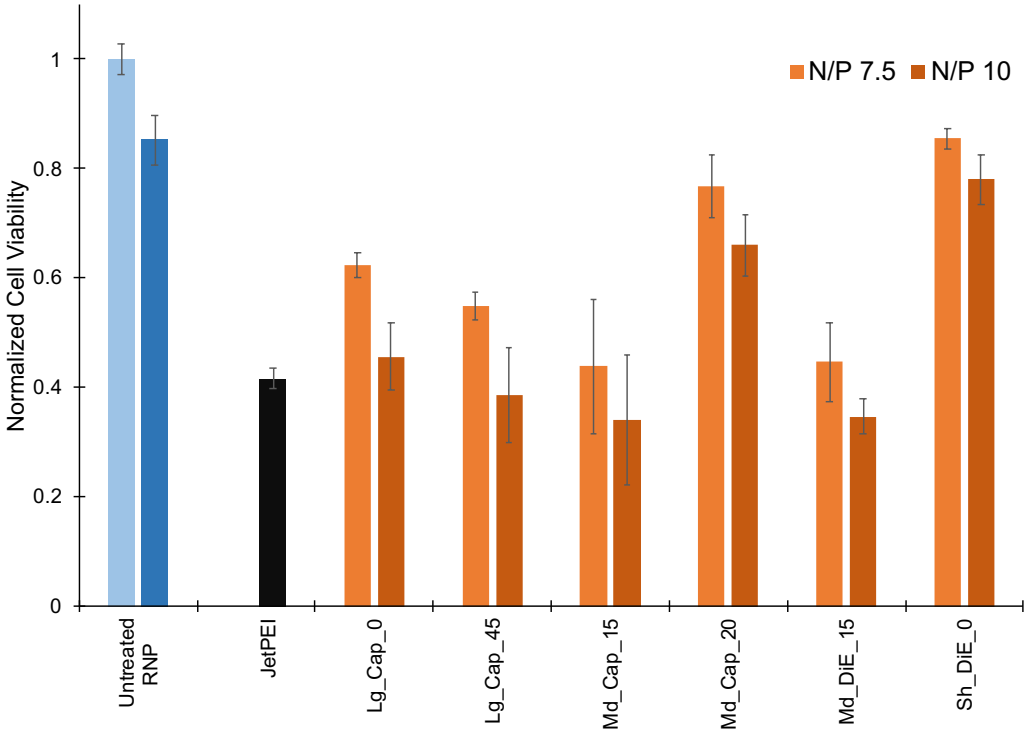


Figure S39. Viability data showing the normalized transmittance in a CCK8 assay in a round 2 loop study with select polymers delivering RNP into HEK293T cells at an N/P ratio of 7.5 and 10.

Round 3 pDNA and mCherry transfection

Based on predictions of the Bayesian machine learning model, eight individual polymer formulations were decided upon for each biological payload. Polyplex formation, transfection, CCK-8 analysis, and flow cytometry workup were all followed as listed above in the general procedures.

For pDNA delivery: Sh_DiE_40 (N/P 11.25, 14.125, 16.375), Lg_Cap_20 (N/P 11), Lg_DiE_40 (N/P 17.5), Lg_Cap_40 (N/P 13.25), Lg_DiE_20 (N/P 13), Md_DiE_40 (N/P 17.125).

For RNP delivery: Lg_Cap_0 (N/P 4, 5.5, 6), Sh_DiE_20 (N/P 8, 10), Lg_Cap_20 (N/P 9.625), Lg_Cap_10 (N/P 8.625), and Lg_DiE_20 (N/P 8.875)

Table S2. Table of %GFP and cell viability data for the 8 polymers transfected with pDNA.

N/P	Polymer	%GFP	Std. Dev.	Viability	Std. Dev.
11.25	Sh_DiE_40	45.6	2.2	0.61	0.08
11	Lg_Cap_20	50.2	1.6	0.16	0.01
14.125	Sh_DiE_40	64.0	3.0	0.42	0.03
17.5	Lg_DiE_40	58.1	0.9	0.44	0.02
16.375	Sh_DiE_40	63.8	2.3	0.42	0.01
13.25	Lg_Cap_40	35.3	2.3	0.45	0.00
13	Lg_DiE_20	29.4	1.7	0.18	0.01
17.175	Md_DiE_40	62.2	1.7	0.49	0.02
	pDNA	0.1	0.0	0.83	0.02
	Untreated	0.0	0.0	1.00	0.07
	JetPEI	76.7	1.8	0.38	0.04

Round 3 mCherry Expression and Cell Viability

Table S3. Table of %mCherry and cell viability data for the 8 polymers transfected with RNP.

N/P	Polymer	%mCherry	Std. Dev.	Viability	Std. Dev.
5.5	Lg_Cap_0	3.54	0.25	0.61	0.08
6	Lg_Cap_0	4.15	0.22	0.16	0.01
10	Sh_DiE_20	1.88	0.25	0.42	0.03
4	Lg_Cap_0	1.74	0.36	0.44	0.02
8	Sh_DiE_20	1.67	0.04	0.42	0.01
9.625	Lg_Cap_20	3.34	0.19	0.45	0.00
8.625	Lg_Cap_10	4.34	0.70	0.18	0.01
8.875	Lg_DiE_20	3.80	0.19	0.49	0.02
	RNP	0.19	0.03	0.83	0.02
	Untreated	0.02	0.02	1.00	0.07
	JetPEI	1.04	0.08	0.38	0.04

Analysis by Machine Learning

Polymer feature attribution using SHAP

The structure-property relationships are explored using the approach explained in Kumar *et al.*[3,4] We fit a second machine learning model called an "explanation model", a scikit-learn[5]extra trees regression model is used to predict the expression and viability for each cargo based on the following polyplex features: scaffold RU, cation type, %BET, pK_a, clogP, polyplex size (R_h), formulation (N/P) ratio, and binding strength. The cation is represented by using an extended connectivity fingerprint[10] and then reducing it to a single value using Principal Component Analysis (PCA). We follow this approach to encoded molecular similarity instead of giving discrete arbitrary values. The encoded values for Cys, Cap, DiE values are 0.42, 0.74 and 0.82 respectively. The model and its hyperparameters were chosen based on satisfactory performance on polymer-grouped 5-fold cross validation. The performance of other common machine learning models is summarized in Table S4. We chose the final extra trees models to perform SHAP analysis.

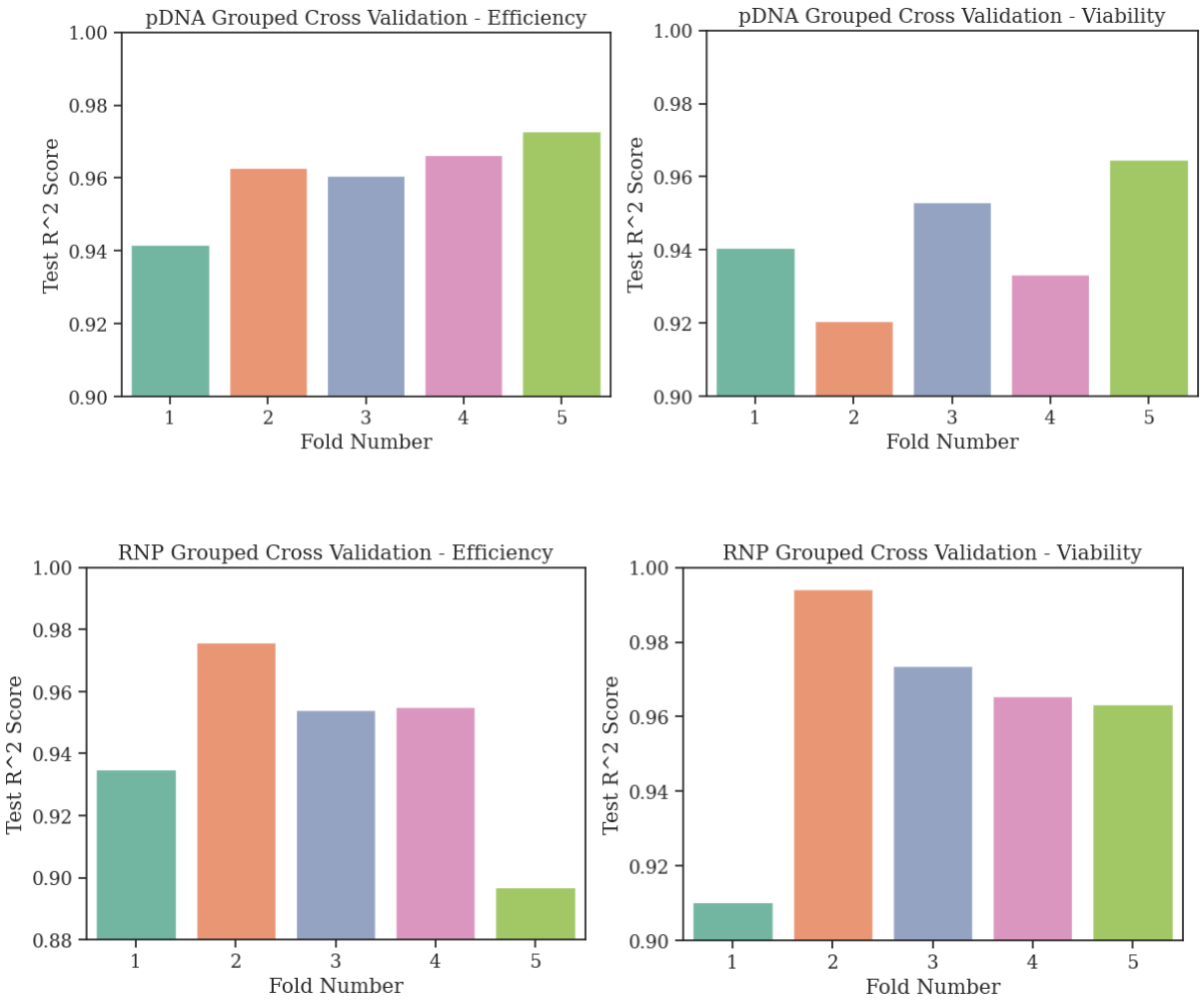


Figure S40. Grouped cross validation results across folds for pDNA and RNP using Extra Trees regression model.

Model	Mean R^2 for efficiency	Mean R^2 for viability	Hyperparameters (Optimal and grid search range)
Extra trees	0.951	0.952	n_estimators= 150 [50, 1000] max_depth = None [2, 7]
Random forest	0.943	0.951	n_estimators= 150 [50, 1000] max_depth = None [2, 7]
Gradient Boosting	0.894	0.911	learning_rate = 1e-2 [1e-4, 0.1] n_estimators = 100 [50, 1000]
Gaussian Process Regressor	0.893	0.916	normalize_y = False [False, True] length_scale_bounds= None [None, "fixed"]

Table S4. Grouped cross validation results for various models, hyperparameter tuning was performed with nested cross validation.

We use this trained model for interpretability via SHapley Adaptive ExPlanations (SHAP).[6] SHAP provides local and global explanations for the relative importance of each feature by learning a local linear model with game-theoretic constraints. In particular, we apply the TreeSHAP algorithm[6] and take the mean absolute SHAP value across all data points as our feature importance (radar plots) and analyze trends in local SHAP values. We evaluated other machine learning models, such as random forests and gradient boosting, and found consistent SHAP value trends.

Bayesian Optimization (BO)

Bayesian optimization (BO) is an additional machine learning approach that employs batch learning to efficiently explore combinatorial design spaces. BO provides a sequential optimization model that identifies discrete predicted choices based on variable design, which can decrease the experimental workload through effective sampling to optimize outputs. With our current scaffolds, when considering multifactorial optimization of pK_a, clogP, %BET, scaffold RU, and formulation N/P ratio, a uniform or quasi-random sampling (e.g., using design of experiments) of the variable space may hide potentially high performing polymer sets and may require a large number of samples to identify the best performing polymers.

Bayesian optimization (BO) is a technique for efficient global optimization of a black-box function.[7] In our case, we are interested in finding the optimal design parameters X^* that maximize our target noisy function, $f(\mathrm{X})$, which corresponds to the measured expression in a particular cargo:

$$\textbf{Equation S4}.\ X^* = \operatorname{argmax} \mathbb{E}[f(\mathrm{X})]$$

For this purpose, we build a probabilistic (Gaussian process[8]) model based on the polymer design parameters: pK_a, clogP, %BET, scaffold RU, and formulation N/P ratio. As the polymers are optimized in a mixed space of continuous (N/P) and discrete variables (Cation, %BET Incorporation, Scaffold RU), we use the mixed space kernel implementation available on the Meta's Ax package.[9] To improve predictive performance of the probabilistic model, we map the cation monomer information to an PCA-encoded representation based on its extended connectivity fingerprint,[10] appended to calculated clogP and pK_a values. This mapping leads to a model that better encodes similarities across monomers compared to purely categorical choices and is thus easier to optimize using Gaussian process regression. We perform three rounds of Bayesian optimization for each cargo. In each case, we synthesize a batch of polymers and measure *in vitro* performance, then we update our Gaussian process model with the expression results and propose a batch of new samples by using an expected improvement acquisition function and Thomson sampling[7], a common choice to balance exploration and exploitation in batch BO. This batch of new samples correspond to a samples that are expected to have good performance by being similar to explored areas of the design space (exploitation) or samples in areas of the design space that have high expected variability (exploration).[9] For each round, we verified that the suggested samples satisfied basic cellular toxicity requirements by fitting the viability output with an auxiliary random forest model. If the predicted viability was under 0.4 for the proposed polymer designs, we removed the sample from the batch, enforcing a basic viability constrain during optimization. The viability threshold was selected empirically based on the initial viability measurement and monitored for an acceptable R^2 (>0.8) in each round.

***In vivo* hydrodynamic tail vein**

All experiments were performed in compliance with the relevant guidelines from the IACUC committee at the University of Minnesota under an approved protocol.

Animals

7-8-week-old C57BL/6NHsd female mice were purchased from Envigo (Madison, WI). Prior to performing study procedures beyond ear tagging and weighing, mice were acclimated to the facility for one week. The mice were housed five per cage and provided food and water ad libitum according to the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)**.** Handling of mice was performed in accordance with the Institutional Animal Care and Use Committee (IACUC) using an Institutional Biosafety Committee (IBC) approved protocol (1908-37343H)

Reagents

D-Luciferin (potassium salt) was purchased from Gold-Bio (St Louis, MO). The luciferin was dissolved in sterile DPBS without calcium or magnesium at a concentration of 15 mg/mL, filter sterilized through a 0.22 μM syringe filter, and stored at -80 °C in black polypropylene tubes.

Hydrodynamic Injections.

The polyplexes were prepared in a dextrose 5% in water (D5W) and at a concentration to deliver 10 μg DNA at an N/P of 5 in 2.0 mL of D5W. The polyplexes were kept on wet ice. The mice were approximately 20 g at the start of the study. Mice were anesthetized with a cocktail of

ketamine hydrochloride (10 mg/kg), acepromazine (0.12 mg/kg), and butorphanol (0.012 mg/kg) by intraperitoneal (IP) injection. Mice were sedated and ataxic but maintained consciousness and a blink reflex as they should for hydrodynamic injections. The 2.0 mL of polyplex/D5W solution was first drawn up into a 3.0 mL syringe using a 20g needle. The needle was then removed and a 27-gauge butterfly catheter needle with an 8-inch line was attached. The mice were warmed under a heat lamp to aid thermoregulation and vasodilation, and the polyplex solution in the syringe was warmed to room temperature. After placing the mice in a restrainer, the catheter needle was inserted into a lateral tail veil and the entire volume of fluid in the syringe was injected into the mouse within 4-5 seconds. After removing the needle, the injection site was held with gauze until hemostasis was achieved. After the procedure, the mice were delivered to a recovery cage and placed on a heating pad set on a low setting. It was empirically determined that there was a 0.3 mL hold-up volume in the catheter tubing, thus the actual amount of vector DNA injected was 8.5 µg/mouse.

***In vivo* bioluminescent imaging**

Imaging was carried out using the IVIS Spectrum *in vivo* imaging system (IVIS, PerkinElmer Inc., Waltham, MA). Briefly, mice were injected IP with 200 µl D-Luciferin, and after 7-8 minutes placed in an induction box for isoflurane anesthesia (5% for induction and 2.5-3% for maintenance). The mice were then transferred to the anesthesia nose cones (at 3% isoflurane) on the imaging platform of the IVIS Spectrum. The mice were imaged 10 minutes post-luciferin injection; the time of maximum signal as determined from preliminary kinetics studies. The IVIS imaging parameters were set for full spectrum bioluminescence, a maximum of 6000 counts or 5 minutes (whichever came first) binning from 2-8, and F/stop from 1-8. Quantitation of bioluminescent signal intensity and normalization of photographic images were carried out using the Living Image software (Perkin Elmer).

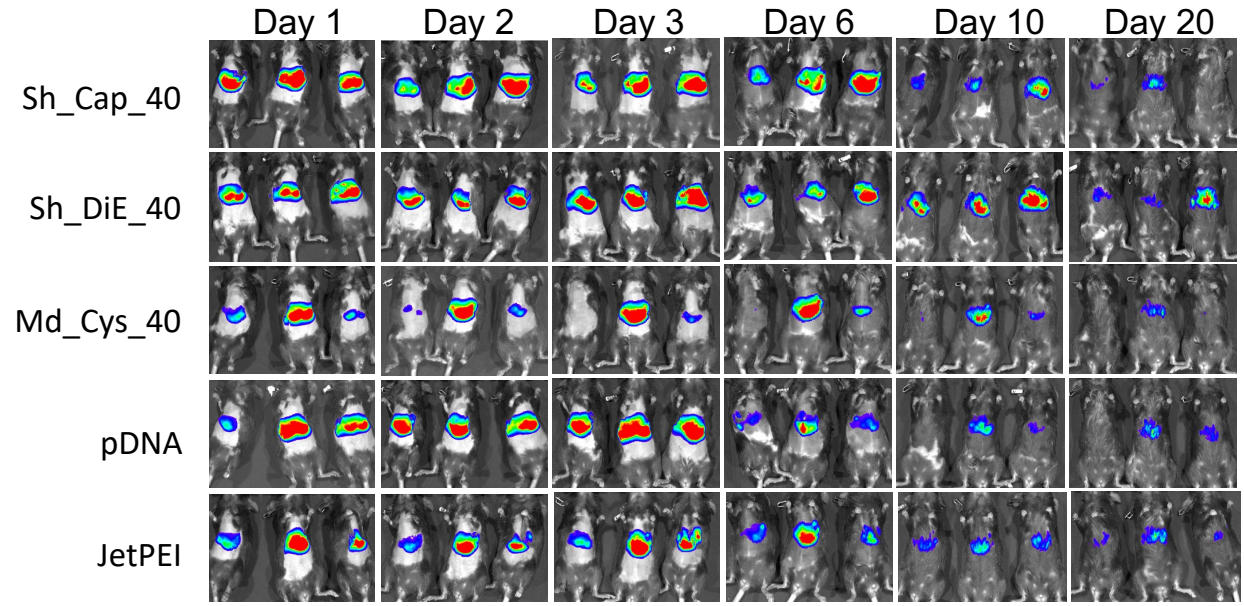


Figure S41. Images of mice in triplicate over the 20-day kinetic *in vivo* experiment showing luminescence.

References

1 L. Jiang, X. Huang, D. Chen, H. Yan, X. Li and X. Du, *Angew. Chemie*, 2017, **129**, 2699–2703.

2 A. M. Striegel, *Chromatographia*, 2017, **80**, 989–996.

3 R. Kumar, N. Le, F. Oviedo, M. E. Brown and M. Theresa, , DOI:10.33774/CHEMRXIV-2021-CDMPF.

4 R. Kumar, N. Le, Z. Tan, M. E. Brown, S. Jiang and T. M. Reineke, *ACS Nano*, 2020, **14**, 17626–17639.

5 Scikit-Learn.

6 R. Sayres, A. Taly, E. Rahimy, K. Blumer, D. Coz, N. Hammel, J. Krause, A. Narayanaswamy, Z. Rastegar, D. Wu, S. Xu, S. Barb, A. Joseph, M. Shumski, J. Smith, A. B. Sood, G. S. Corrado, L. Peng and D. R. Webster, *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp 4768−4777.

7 B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. de Freitas, *Proc. IEEE*, 2016, **104**, 148–175.

8 C. E. Rasmussen, in *Zhurnal mikrobiologii, epidemiologii, i immunobiologii*, 2004, vol. 30, pp. 63–71.

9 facebook/Ax.

10 D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742–754.